Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 23 (2021), Issue 3 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Viewpoints

Original Papers

Letter to the Editor

Corrigenda and Addendas

Review

The Effect of Information and Communication Technology and Social Networking Site Use on Older People's Well-Being in Relation to Loneliness: Review of Experimental Studies

Georgia Casanova[1], MSc; Daniele Zaccaria[2], PhD; Elena Rolandi[3], MSc; Antonio Guaita[3], MD

[1]Centre for Socio-Economic Research on Ageing, National Institute of Health & Science on Ageing, Istituto di Ricovero e Cura a Carattere Scientifico, Ancona, Italy

[2]Centre of Competence on Ageing, Department of Business Economics, Health and Social Care, University of Applied Sciences and Arts of Southern Switzerland, Manno, Switzerland

[3]Golgi Cenci Foundation, Abbiategrasso, Italy

Corresponding Author:
Georgia Casanova, MSc
Centre for Socio-Economic Research on Ageing
National Institute of Health & Science on Ageing
Istituto di Ricovero e Cura a Carattere Scientifico
Via Santa Margherita 5
Ancona
Italy
Phone: 39 347 0836007
Email: g.casanova@inrca.it

Abstract

Background: In the last decades, the relationship between social networking sites (SNSs) and older people's loneliness is gaining specific relevance. Studies in this field are often based on qualitative methods to study in-depth self-perceived issues, including loneliness and well-being, or quantitative surveys to report the links between information and communication technologies (ICTs) and older people's well-being or loneliness. However, these nonexperimental methods are unable to deeply analyze the causal relationship. Moreover, the research on older people's SNS use is still scant, especially regarding its impact on health and well-being. In recent years, the existing review studies have separately focused their attention on loneliness and social isolation of older people or on the use of ICTs and SNSs in elderly populations without addressing the relationship between the former and the latter. This thorough qualitative review provides an analysis of research performed using an experimental or quasi-experimental design that investigates the causal effect of ICT and SNS use on elderly people's well-being related to loneliness.

Objective: The aims of this review are to contrast and compare research designs (sampling and recruitment, evaluation tools, interventions) and the findings of these studies and highlight their limitations.

Methods: Using an approach that integrates the methodological framework for scoping studies and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for systematic reviews, we identified 11 articles that met our inclusion criteria. A thematic and content analysis was performed based on the ex post categorization of the data on the selected studies, and the data were summarized in tables.

Results: The analysis of the selected articles showed that: (1) ICT use is positively but weakly related to the different measures of older people's well-being and loneliness, (2) overall, the studies under review lack a sound experimental design, (3) the main limitations of these studies lie in the lack of rigor in the sampling method and in the recruitment strategy.

Conclusions: The analysis of the reviewed studies confirms the existence of a beneficial effect of ICT use on the well-being of older people in terms of reduced loneliness. However, the causal relationship is often found to be weak. This review highlights the need to study these issues further with adequate methodological rigor.

(J Med Internet Res 2021;23(3):e23588) doi:10.2196/23588

KEYWORDS

review; aging; loneliness; older people's well-being; ICTs; social network sites

Introduction

The percentage of people aged 65 years and older in the world will rise to 24% in 2030 [1]. Literature underlines how loneliness is one of the main risk factors that have negative effects on seniors' health [2]. Loneliness is usually defined as an undesirable subjective experience related to unfulfilled intimate and social needs [3]. In Europe, between 10% and 20% of elderly people in Western and Northern Europe and 30% to 55% in Eastern Europe declared feeling lonely [4]. Many interventions have been adopted to reduce loneliness and increase the well-being of elders. Among them, information and communication technologies (ICTs) have been used to help older adults cope with loneliness [5]. The use of ICTs has grown significantly since the 2000s [6]. ICTs are those technologies that can be used to interlink information technology devices such as PCs with communication technologies such as telephones and their telecommunication networks. Michiels and Van Crowder [7] reported almost 20 years ago how ICTs are an expanding assembly of technologies. In the last decades, PCs, laptops, smartphones, tablets with email, and the internet are examples of ICTs able to connect people and support their social life [8]. These types of ICTs, intended to alleviate loneliness and social isolation among older people, are considered significant in expanding and sustaining social contact and improving emotional well-being [9] and are the focus of this review. Recently, experts' attention has also been drawn to the role of technologies aimed at promoting social relationships, such as social networking sites (SNSs) [10]. Boyd et al [11] defined the SNSs as web-based services that allow individuals to (1) construct a public or semipublic profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. At the moment, Facebook, Instagram, LinkedIn, and Twitter are the most popular SNSs. In particular, Facebook seems to be the most used by older people [12]. The relationship between SNSs and loneliness in older people is gaining specific relevance. However, research on older people's SNS use is still scant, especially regarding their impact on health and well-being. Eggermont and colleagues [13] report that older people consider SNSs useful tools to contrast loneliness, one that should integrate (but not replace) face-to-face contacts. Gibson et al [14] have found that older adults are concerned about privacy issues. Some studies also examine older people's characteristics that favor SNS adoption. For example, Liu et al [15] report that, in the United States, elderly users of SNSs are more likely to be younger, female, and widowed. Studies in this field are often based on qualitative methods, because they allow studying in-depth self-perceived issues, including loneliness and well-being, but to a lesser extent the causal relationship between them [16]. Among quantitative studies, the survey analysis is most used to report the links between ICTs and older people's well-being [17] or loneliness [18]. Despite the increased attention gained by these studies, the survey—often applied to cross-sectional analysis—does not allow the detection of a causal relationship. In 2013, Nef et al [12] underlined a shortage of experimental research on the relationship between ICT use and older adult well-being, considered more adequate to analyze the causal relationship between phenomena [19].

In recent years, review studies have focused on loneliness and social isolation of older people [20,21] or on the use of ICTs and SNSs in elderly populations without tackling the relationship between the former issue and the latter [12,22]. One of the few literature reviews focused on the relationship between ICT use and loneliness in older people underlines how this research field involves different theoretical frames from various scientific fields [22]. The existing quantitative experimental studies on these issues have significant limitations that may hamper the relevance of the research findings, such as small and not representative samples [23].

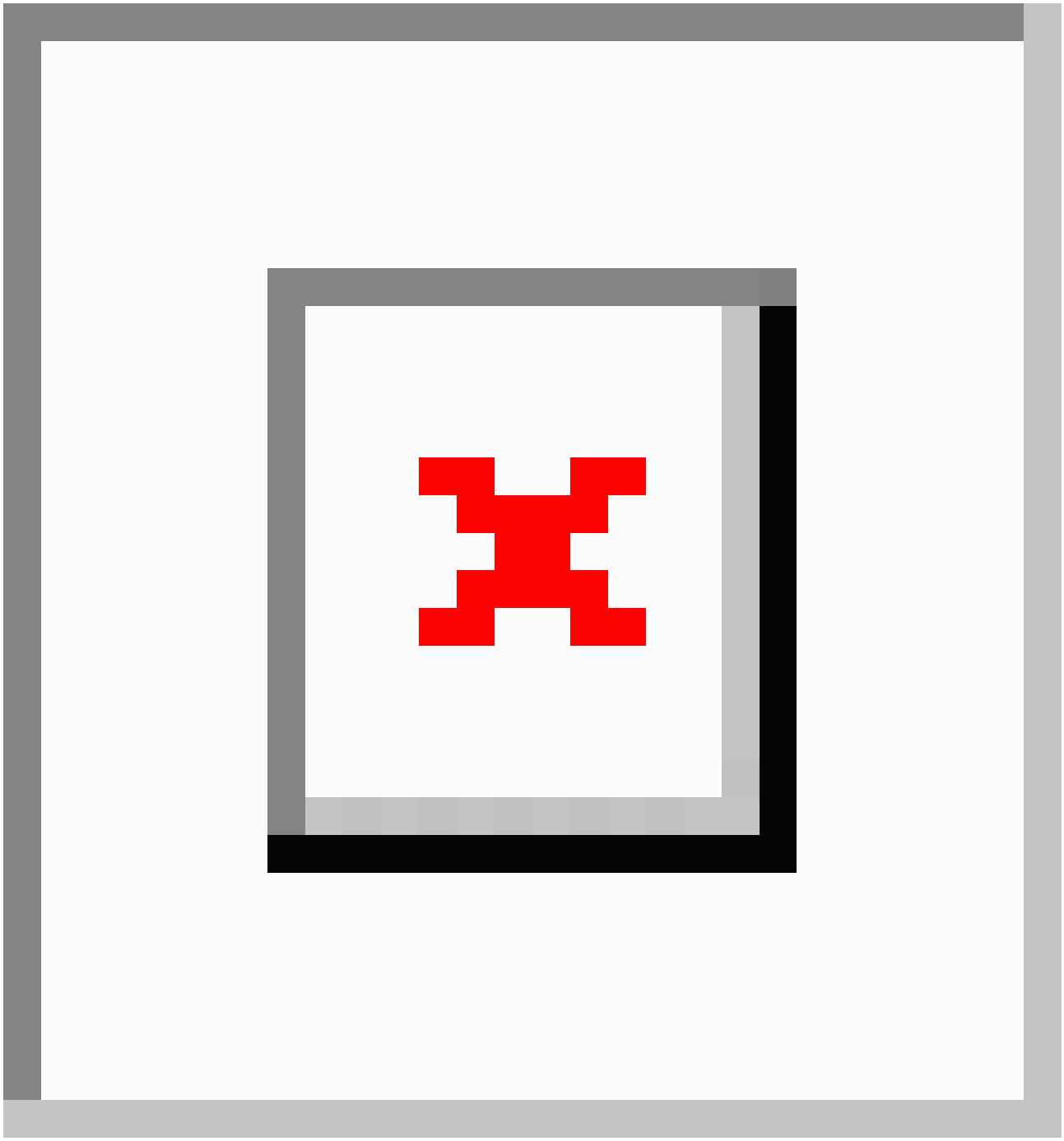
The purpose of this review is to contribute to the literature debate on the effect of SNS use on older people's well-being with specific attention on loneliness, focusing on experimental and quasi-experimental studies. The main aims are to examine and compare the selected studies, analyzing their protocols (sampling, evaluation tools, and treatments) and their findings to identify strengths and limitations and support the development of further similar studies. According to the aims, we present a thorough qualitative review of experimental studies in this field.

Methods

Review Procedures

To ensure a high quality of reporting, this study used the 5 stages for reviews proposed by Arksey and O'Malley [24]: (1) identification of research questions; (2) identification of relevant studies; (3) selection of studies; (4) charting the data; and (5) collating, summarizing, and reporting the results. Taking into account the low spread of the analyzed studies, we decided to improve the method path by the integration of stages 2 and 3 with the 4 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) stages: identification, screening, eligibility, and inclusion [25]. Figure 1 shows the complete method implemented in this review. The combination of these methods ensures the review will stay linear and focused, as proposed by Arksey and O'Malley [24], and limits losing useful papers on the topic thanks to PRISMA approach.

Figure 1. Flowchart of scoping review: methods and procedure.



Identification of Research Questions and Selection Process

In the first stage, we identified 3 research questions: (1) What are the main methodological characteristics of the study designs? (2) What causal evidence exists on the relationships between ICTs, SNSs, and well-being in older people, particularly as regards loneliness? (3) What are the main studies' declared limitations?

In March 2020 for stages 2 and 3, we performed a comprehensive literature search in the main search engines used in health and social sciences: Scopus, PubMed, Web of Science, and Sociological Abstracts. The search was based on a set of keywords ("older people," "elderly," "loneliness," "well-being," "information and communication technologies," "social networks," and "Facebook") clustered in the search process to enhance the relationship between web-based communication technologies, isolation, and loneliness conditions of older people, as detailed in Figure 1. Facebook has been included in the set of keywords because, as specified in the introduction, this SNS is most used by older people. The search was conducted in English. We found 204 records, without duplicates, published from 2002 to 2019, that met the following eligibility criteria: (1) study uses experimental or quasi-experimental design; (2) study is based on pre-post controlled trials; (3) loneliness and/or well-being are outcomes; and (4) the target population is people aged over 60 years, considered the threshold for the aging process [26]. Two independent

researchers (GC, DZ) screened the identified studies for their relevance based on title and abstract, and 185 studies were excluded: records not pertinent (n=69); qualitative studies (n=67); loneliness and well-being were not considered as the outcome (n=23); methodological papers (n= 6); literature reviews (n=13); and target population includes people under 60 years (n=7). A total of 19 articles met the inclusion criteria. Discrepancies between data extractions were resolved by team discussion and involving a third person (AG) as an arbitrator. After carefully reading the full texts of the articles, researchers excluded 8 more studies because they did not adopt experimental or quasi-experimental design. The selected articles were retrieved in full text and reevaluated by the authors for the final consensus on the inclusion. In the end, 11 papers were included in this review. No other references were identified by hand searching or analyzing the references of included articles.

Data Extraction, Data Synthesis, and Analysis

In stage 4, we organized the materials to be analyzed. First, we ordered the collected papers by date from the oldest to the newest [27-37]. Moreover, we identified the review's framework and related categories to provide our analysis (detailed in Table 1). Two macro areas of analysis have been identified: main characteristics of the protocols and main contents of the studies. For both of them, we identified a set of specific categories of analysis, one or more related main questions, and the items to collect. In stage 5, two researchers (GC, DZ) independently extracted the items based on the identified categories. To collect similar information on all studies, we performed a thematic and content analysis [38] based on the ex-post categorization of variables (eg, specific items) [39] to (1) detect the presence of variables in each selected study, (2) identify different modalities of selected variables (eg, tools used or different choices on sampling process), and (3) make them easy to read based on classification and summarization of specific contents. Last, to answer our research questions, the first author completed the data analysis using standardized self-made forms corresponding to Tables 2-4.

Table 1. Frameworks of analysis: general items, main questions, and detected items by analysis macro areas.

Categories and general items	Main questions	Detected items
Protocol characteristics		
Keywords	• What are the items of study declared by authors?	• List of keywords
Methods	• Is the study experimental?	• Presence of randomized trial study
Sample	• What is the dimension of the sample?	• Number of participants in assessments
Population	• What are the age characteristics of population?	• Range of age group • Mean age
Selection criteria	• What are the applied selection criteria?	• Place of living • Previous experience with ICT[a] use • Healthy conditions • Level of social engagement
Recruitment	• What is the recruitment strategy?	• Place of recruitment
Control group	• How many control groups are included in the protocol?	• Presence of control group • Typologies of control groups
Assessments	• How many assessments are included?	• Number and timing of follow-up evaluations
Intervention	• How long is the intervention? • What activities are performed in the intervention?	• Number of intervention weeks or years • Presence of training course and training size
Contents of studies		
Study issues	• Is the study focused on SNS[b] use?	• Presence of SNS focus
Outcomes	• What are the outcomes included in the study? • Is loneliness one of them?	• List of outcomes • Presence of loneliness as an outcome
Evaluation tools	• What are the evaluation tools used in the study? • Which validated scales have been used?	• Presence of validated scales • Presence of self-made instruments • List of validated scales
Main results	• What are the main results of study?	• Summarize the results
Declared limitations	• What are the main limitations of study?	• Summarize the limitation

[a]ICT: information and communication technology.

[b]SNS: social networking site.

Table 2. Keywords and protocol characteristics.

Author name and year	Keywords	Randomized trial	Sample size, baseline (f/u[a])	Age range (mean)	Recruitment in care home	Control group	>1 f/u	Length of training (intervention)
White et al (2002) [27]	Older people; social isolation; internet; psychosocial impact	*[b]	100 (84)	59-83 (71.5)	*	P[c]	no	9 h (20 w)
Fokkema & Knipscheer (2007) [28]	N/A[d]	no	21 (12)	66+ (N/A)	no	V[e]	*	10 h (3 yrs)
Shapira et al (2007) [29]	Internet; senior well-being; personal sense of empowerment; Israel	no	46 (39)	70-93 (81.2)	*	A[f]	no	20 h (15 w)
Siegers et al (2008) [30]	Computer use; internet; well-being	*	236 (211)	64-75 (N/A)	no	P,P	*	4 h (54 w)
Woodward et al (2010) [31]	Gerontology; information and communication technologies; older adults; computer training; social support; mental health	*	83 (83)	60-89 (71.8)	no	P	*	24 h (24 w)
Blažun et al (2012) [32]	Older people; loneliness; computer training course; socialization; health; well-being	no	58 (45)	58-93 (72.9)	*	no	no	4 h (3 w)
Cotten et al (2012) [33]	Computers; internet; loneliness; social isolation; older adults; independent living; assistant living facilities	*	205 (205)	N/A (82.7)	*	A,P	*	N/A (8 w)
Myhre et al (2017) [34]	Executive functions; social interaction; social media; technology; training; working memory	*	43 (41)	75-86 (81.7 / 75.7?)	*	A,P	no	6 h (8 w)
Larsson et al (2016) [35]	Loneliness; social activities; social contact	*	30 (30)	61-89 (71.2)	no	P	*	N/A
Quinn (2018) [36]	Older adults; executive functions; social media training; experiment	*	34	N/A (76.5)	no	P	*	8 h (4 w)
Morton et al (2018) [37]	computers; internet; social connections; cognitive capacity; well-being	*	97 (76)	60-95 (80.7)	* not exclusive	P,D[g]	no	18 h (12 w)

[a]f/u: follow-up.

[a]*: yes, not available, or not declared in paper.

[a]P: passive control group or waiting list.

[a]N/A: not applicable.

[a]V: virtual control group by online survey.

[a]A: active control group.

[a]D: double intervention with different place of living and two control groups.

Table 3. Contents of studies: focus on social networking sites, outcomes, and tools.

Author name and year	Focus on SNS[a]	Loneliness	Beneficial effects on loneliness	Other outcomes	Specific questionnaire	Validated scales
White et al (2002) [27]	no	*[b]	no	Psychological and social well-being	no	UCLA Loneliness Scale, UCLAS[c], CES-Depression[d], PCLS[e], CAS[f]
Fokkema and Knipscheer (2007) [28]	no	*	*	Distinction between social loneliness and emotional loneliness	no[h]	SJGLS-6[g]
Shapira et al (2007) [29]	no	*	*	Psychological well-being	no	UCLAS, DPFS[h], DACL[i], SAS[j], PCS[k], LSS[l]
Siegers et al (2008) [30]	no	*	no	Physical, social, and emotional well-being	no	SJGLS-6, SF-36[m], SCL-90[n], IADL scale[o], EPQ-R[p], ECS[q]
Woodward et al (2010) [31]	no	*	no	Mental health	*	Antonucci's HMT[r], Gagnè-MNSS[s], CSE-16[t]
Blažun et al (2012) [32]	no	*	*	no	*	no
Cotten et al (2012) [33]	*	*	* weakly	Social well-being	*	RTLS-34[u]
Myhre et al (2017) [34]	*	*	no	Cognitive functions	no	UCLA Loneliness Scale, LSNS-18[v], SPS-10[w], RAVLT[x], ReyCFT[y], DSST[z], DFRTT[aa], TMT[bb], COWAT[cc], Miyake EFsT[dd]
Larsson et al (2016) [35]	*	*	*	no	no	UCLA Loneliness Scale, ESI[ee], VAS[ff]
Quinn (2018) [36]	*	no	no	Cognitive functions	no	MMSE[gg], COAST[hh], SDMT[ii], WAIS[jj]
Morton et al (2018) [37]	*	*	no	Computer attitude; sense of self-worth (competence, autonomy, and personal identity); cognitive and mental health	no	UCLA Loneliness Scale, SNAI[kk], ACE-R[ll], CES-D Depression[mm], GAI-SF[nn], GHQ-12[oo], SWL[pp], SWLS[qq], CAS

[a]SNS: social networking site.
[b]*yes or not available.
[c]UCLAS: Revised UCLA Loneliness Scale.
[d]CES-Depression: CES-Depression Scale.
[e]PCLS: Perceived Control Life Situation.
[f]CAS: Computer Attitude Scale.
[g]SJGLS-6: 6-Item De Jong Gierveld Loneliness Scales.
[h]DPFS: Difficulties in Physical Functioning Scale.
[i]DACL: Depressive Adjective Checklist.
[j]SAS: Self-Anchoring Scale.
[k]PCS: Perceived Control scale.
[l]LSS: Life Satisfaction Scale.
[m]SF-36: 36-item Short Form Health Survey.
[n]SCL-90: 90-item Symptom Check List.
[o]IADL scale: Specific Questionnaire to Measure Daily Activities.
[p]EPQ-R: Subscales of the Eysenck Personality Questionnaire.
[q]ECS: External Control Scale.
[r]Antonucci's HMT: Antonucci's Hierarchical Mapping Technique.
[s]Gagnè-MNSS: Gagnè Motivation and Need Satisfaction Scale.
[t]CSE-16: Computer Self-Efficacy 16-item Scales.
[u]RTLS-34: Rasch Type Loneliness Scale 34-item.
[v]LSNS-18: Lubben Social Network Scale 18-item version.

[w]SPS-10: Social Provision Scale.
[x]RAVLT: Rey Auditory Verbal Learning test.
[y]ReyCFT: Rey Complex Figure Test.
[z]DSST: Digit Symbol Substitution Test.
[aa]DFRTT: Deary-Liewald Reaction Time Test.
[bb]TMT: Trail Making Test.
[cc]COWAT: Controlled Oral Word Association Test and Category Fluency Test.
[dd]Miyake EFsT: Miyake Executive Function Test.
[ee]ESI: Evaluation of Social Interaction.
[ff]VAS: visual analog scale.
[gg]MMSE: Mini-Mental State Examination.
[hh]COAST: California Older Adults Stroop Test.
[ii]SDMT: Symbol Digit Modalities Test.
[jj]WAIS: Wechsler Digit Span Forward and Backward subtest.
[kk]SNAI: Social Networking Activity Index.
[ll]ACE-R: Addenbrooke's Cognitive Examination–Revised.
[mm]CES-D Depression: CES-D Depression Scale.
[nn]GAI-SF: Geriatric Anxiety Inventory Short Form.
[oo]GHQ-12: General Health Questionnaire.
[pp]SWL: 5-Item Satisfaction With Life Scale.
[qq]SWLS: The Satisfaction With Life Scale.

Table 4. Contents of studies: findings and limitations.

Author name and year	Results	Limitations
White et al (2002) [27]	• Successful intervention. Improvement of PC use by all participants • No significant correlation between self-reported health, activity limitations, and internet use	• Short follow-up time to evaluate intervention effect in terms of loneliness • Need for more intensive intervention • Use of self-reported condition of PC use for selection criteria
Fokkema and Knipscheer (2007) [28]	• Feeling of loneliness significantly decreased in both follow-ups (2 and 3 years after baseline) • Greatest reduction took place in the first step (T0 to T1) • Difference of reduction of loneliness between intervention group and control group is significant • Emotional loneliness reduction is significant, unlike the results on social loneliness • Qualitative findings confirm results that the internet helps to maintain contact with family or friends and it is a meaningful way to pass time	• Social contacts of participants increased due to classes and tutoring included in the intervention. These could influence effects of loneliness • Experiment was only performed once, and results should then be checked by repeat study • Suboptimal matching of intervention and control groups created a problem with identifying the intervention as the reason for the reduction of loneliness
Shapira et al (2007) [29]	• Positive effects on loneliness, less depression, more satisfaction with life, and more control and pleasure with their current quality of life	• Short and simple intervention to analyze a complex phenomenon • Language barrier in PC use • Lack of analysis of secondary effects (eg, social environment and activism)
Siegers et al (2008) [30]	• No significant differences (including for loneliness) were observed between groups and between times of follow-ups	• Lack of analysis of changes in lifestyles • Self-report measures of aspects of well-being • Exposure time bias: 1 year too short to make structural changes but too long to measure small daily changes in attitude
Woodward et al (2010) [31]	• No significant effect on loneliness and depressive symptoms between groups and between times • Experimental group reported a significantly higher quality of life compared with control group • No significant effect on social support outcome • Significantly greater computer self-efficacy and use of ICTs[a] for both groups	• Randomization on people who agreed to participate made a self-selected sample not corresponding to a real population
Blažun et al (2012) [32]	• Loneliness reduction was statistically significant, detected in pre-post analysis related to gender (female), living alone, living in town • Improvement in independence of people living alone in town and their perception of safety	• Different trials in country cases • Data collection bias; questionnaire translation
Cotten et al (2012) [33]	• Weak and negative correlation between going online and loneliness • Moderate correlation between internet outcome variables and quality of communication • Mean social relation network: 11.2 members—friends and family	• Small sample • Lack of analysis of detailed information on participants
Myhre et al (2017) [34]	• Social outcome measured no significant change between pre- and posttimes	• Small sample • Randomization sample had been adapted by availability of individuals • Socialization in group should have had an effect on results
Larsson et al (2013) [41]	• Significant effect on satisfaction loneliness • Online contacts were significantly improved • No changes in offline contacts satisfaction	• Lack of active control groups
Quinn (2018) [36]	• No significant differences pre-post • No significant improvement on MMSE[b]	• Small sample • Lack of detailed information on participants

Author name and year	Results	Limitations
Morton et al (2018) [37]	• Significant cognitive improvements across time in the training but not control group • This effect was mediated through a combination of increased social activity, improved self-competence, and preserved personal identity strength • Indirect effects on mental health outcomes via these processes were also observed • Larger improvement was detected in the residential care group (more socially restricted at the outset) • No significant effect on satisfaction loneliness	• Inability to disentangle the effects of training from the trainer visits • High attrition rate • Enduring relevance of the issues raised: importance of providing older adults with supportive training in the social use of internet technology might wane with current generations

[a]ICT: information and communication technology.

[b]MMSE: Mini-Mental State Examination.

Results

Protocol Characteristics: Methods

The main findings are summarized in Tables 2-4. Some methodological differences have been detected in study designs. First, 9 studies [27,29,31-37] were experimental with randomized sampling, while 2 [28,30] were quasi-experimental studies. In addition, 2 studies [27,34] were based on previous pilot studies. In the end, Larsson et al [35] used a crossover study and White et al [27] integrated the quantitative study with a short qualitative interview performed at the follow-up.

Protocol Characteristics: Population, Sample, and Inclusion Criteria

A total of 953 participants were recruited for these studies, and 860 of them took part in the postintervention assessment. Thus, around 10% of participants dropped out of the experimental studies before the follow-up. The age groups were different in each of the analyzed studies: from the largest, 58 to 93 years, in the Blažun et al [32] and Morton et al [37] studies to the smallest, 64 to 75 years [29] and 75 to 86 years [32], defined by 11 years' range (Table 2). The participants mean age, when declared, was over 71 years, with 4 studies ([29,33,34,37] reporting a mean age over 81 years.

Furthermore, the sample sizes were very heterogeneous among studies. Taking into consideration the follow-ups, most of the studies involved fewer than 50 participants [28,29,32,34-36], while in 3 studies, the sample size was more than 75 but less than 85 [27,31,37]. Only Siegers et al [30] and Cotten et al [33] proposed samples comprising more than 200 subjects. Power analysis to support the sample size definition was conducted in only one study [30].

The studies imposed a range of inclusion criteria (20 in total, see Multimedia Appendix 1). In the summarizing phase, we identified 4 categories of criteria: (1) health conditions, (2) level of experience on PC and ICT use, (3) place of residence, and (4) social engagement. Health condition was the most used inclusion criteria: only 4 studies did not provide health condition criteria to take part in the study [28,31,33,35]. Health was frequently conceptualized as cognitive ability. This information was collected as self-reported information by participants [28,34] or caregivers [27,32]. Four studies directly measured the cognitive level using the Mini-Mental State Examination , but heterogeneous thresholds are used to identify cognitively intact individuals: in the Morton et al [37] study 19 points were sufficient to be included in the sample, Siegers et al [30] fixed the minimum score at 24 points, and Blažun et al [32] and Myhre et al [34] put it at 26 points. In most of the selected studies, low level of knowledge and use of technologies has been used as a specific inclusion criterion. Some specific differences were detected: White et al [27], Fokkema and Knipscheer [28], and Morton et al [37] generally focused only on PC use, while Quinn [36] shifted the focus to SNS use. Myhre et al [34] considered both, underlining the complexity of social interaction based on use of technologies. Conversely, Siegers et al [30] and Larsson et al [35] selected people who were already experienced in using a PC. Last, the availability of a PC at home was required by Larsson et al [35] and Myhre et al [34], while Morton et al [37] required available space and infrastructure for internet use.

The place of residence criteria were strongly related to the recruitment process. Five studies [27,29,32-34] involved people in residential homes, communities, or day centers, but only Cotten et al [33] declared that living in residential homes was one of the selection criterion. Morton et al [37] included people living at home and in residential care to build a double comparison sample: intervention and control groups living in 2 different places. The remaining 5 studies recruited exclusively older people living at home.

Social engagement criteria were adopted in 2 studies: Larsson et al [35] selected retired people who reported loneliness and social isolation experiences, while Fokkema and Knipscheer [28] combined reported loneliness experiences with willingness to take part in the study.

Protocol Characteristics: Control Groups, Follow-Up, and Interventions

All studies except for Blažun et al [32] benefited from at least one control group. The waiting list as a passive control group was a choice in 4 papers [27,31,35,36]. Fokkema and Knipscheer [28] chose to use an online survey as a virtual passive control group. Shapira et al [29] included only an active control group, with an alternative intervention to a selected part of the sample. Instead, to support the comparative control action, Myhre et al [34] and Cotten et al [33] combined active and passive control groups. Two passive control groups were used by Siegers et al [30] to compare the results between individuals not interested in attending the training course on PC use and those included

in the waiting list despite their interest. Morton et al [37] identified 2 intervention groups and 2 waiting lists to further detect differences due to living places (at home or residential care institution).

Five protocols were characterized by a single follow-up at the end of the intervention [27,29,32,34,37]. In 2 cases, an additional follow-up was planned at 4 months [36] or 12 [30] months after the intervention. Fokkema and Knipscheer [28] proposed 2 long-term follow-ups at 14 months and 24 months after the baseline. Cotten et al [33] included multiple follow-ups at 3 months, 6 months, and 12 months. This choice is quite similar to what proposed by Woodward et al [31], based on 3 repeat measurements every 3 months after baseline (at 3 months, 6 months, and 9 months). In the crossover study by Larsson et al [35], participants were tested 2 times after baseline, as expected in this kind of method.

All experimental studies used training classes on PC or SNS as the main part of the intervention. In 4 studies, training did not last longer than 9 hours, and the intervention duration was often less than 8 weeks [32-34,36]. Most interventions included the provision of extra incentives to support ICT use (eg, tutoring and exercise sections [27-31,37]). Siegers et al [30] and Fokkema an Knipscheer [28] made a choice to invest in long-term online tutoring. Both offered short training (4 hours and 10 hours) before their long intervention. Siegers et al [30] preferred to use online tutoring, while in the study by Fokkema and Knipscheer [28] participants were supported and coached by visiting volunteers once every 2 or 3 weeks throughout the 3 years of the project. The protocol by Morton et al [37] provided 18 hours of training in the first month of intervention, while the second and third months were devoted to online tutoring.

Contents of Studies: Focal Issues, Outcomes, and Tools Used

Papers published after 2013 considered SNS use an independent variable, while the previous ones focused their attention on PC use, including email activity (Table 3). Loneliness was analyzed as a specific well-being outcome by all studies except Quinn [36], who focused on self-perceived SNS impact on personal social life and personal relationship network without direct reference to loneliness. In particular, loneliness was the single outcome analyzed in Larsson et al [35] and Blažun et al [32]. Fokkema and Knipscheer [28], taking inspiration from Weiss's theory, distinguished between social loneliness and emotional loneliness [40] and provided a measurement of both. Other studies presented additional outcomes focused on the health and well-being of older people: cognitive functions [32,34], cognitive and mental health [31,37], psychological effects [27,29,30], and emotional well-being [30]. The study by Morton et al [37] stressed sense of self-worth meant as autonomy, personal competence and personal identity. Social well-being (meant also as social isolation) was analyzed as participation in social activities [30] or social network ties and measured as quantity and quality of online communication [33,37].

International validated scales were the most commonly used measurement tools, but 2 studies included bespoke questionnaires created by the authors [32,33]. The 44 scales and/or tests were summarized in 5 categories: (1) aspects of social relationship life, (2) neuropsychological conditions, (3) clinical and physical well-being, (4) psychological well-being, and (5) ICT attitude and use. The complete list of validated scales with breakdown by category is provided in Multimedia Appendix 2.

Three studies [27,29,33] used the UCLA Loneliness Scale as a single perceived social measurement tool, while 2 studies combined that scale with others: Myhre et al [34] added the Lubben Social Network Scale 18-item and the Social Provisions Scale and Larsson et al [35] added the Evaluation of Social Interaction scale. Siegers et al [30] and Fokkema and Knipscheer [28] used the RTLS-34 scale exclusively. Woodward et al [31] combined the RTLS-34 with Antonucci's Hierarchical Mapping Technique.

Clinical and physical well-being was investigated by Shapira et al [29] using the Difficulties in Physical Functioning Scale and by Siegers et al [30] using the Lawton Instrumental Activities of Daily Living Scale, 90-item Symptom Checklist ,and the 36-item Short Form Health Survey. Myhre et al [34], Larsson et al [35], and Quinn [36] included neuropsychological tests, measuring intelligence, attention, memory and executive function abilities (as detailed in Table 3). White et al [27], Morton et al [37], Woodward et al [31], Shapira et al [29], and Siegers et al [30] preferred to study the effects on the psychological well-being of older people using the Life Satisfaction Scale, depression level, and mastering of life phenomena. In addition, White et al [27], Morton et al [37], and Woodward et al [31] choose to use the Computer Attitude Scale and the 16-item Computer Self-Efficacy Scale.

Contents of Studies: Findings and Declared Limitations

Five studies highlight the positive effect of ICT use including SNS use on the social relationships of older people (Table 4). The investigations by Shapira et al [29], Fokkema and Knipscheer [28], Blažun et al [32], and Woodward et al [31] highlight beneficial effects on loneliness and depressive symptoms, increased life satisfaction, and more control and pleasure with their current quality of life. One of them, Blažun et al [32], showed the reduction of loneliness stratified by gender: the improvement in perceived loneliness was more common among women than men. Conversely, Siegers et al [30], Morton et al [37], and Myhre et al [34] observed no statistically significant differences in pre-post evaluations, despite the data showing a decreasing trend of loneliness among participants. Even if Morton et al [37] underlined an improvement in both intervention groups included in their study, albeit one not statistically significant, the larger upgrade was detected in people living in residential care compared with those living at home. Fokkema and Knipscheer [28] stressed how the long-term (2 and 3 years after baseline) improvement on loneliness was statistically significant, unlike what had been detected at the first follow-up. However, participants in the training course of ICTs and SNSs improved their competence and sense of self-worth through increased social activity [37], but these positive effects on online communications do not impact personal satisfaction related to offline contacts [35]. Cotten et al [33] confirmed that the improvement of online

activities did not influence the quality of communications or the size of the personal network in the social relationship: if, generally, the network comprised 11 people, most of them were old friends and family members. This assumption was further confirmed by the findings of Fokkema and Knipscheer [28]. The reduction of emotional loneliness proved significant, unlike the results on social loneliness. Moreover, the qualitative findings underline that, for the interviewed participants, the internet was a meaningful way to pass time because it helped maintain contact with family or friends.

In addition to general results on the main declared outcomes, some papers reported the effects of treatment on secondary outcomes in terms of the success of training [27]. Morton et al [37] identified additional indirect effects on mental health outcomes of older people involved in the study. In particular, increased personal competence on ICT use was related to improvement in the sense of self-worth, strength of personal identity, and self-esteem. Similarly, Blažun et al [32] detected an enhanced independence of participants living alone in town and in their perception of safety. The study by Quinn [36] detected the lack of statistical significance of SNS use on cognitive functions in older people.

The authors highlighted how their study protocols presented limitations. In 2 cases, the intervention duration was considered too short to assess ICT effect on older people's loneliness [27,29]. White et al [27] and Siegers et al [30] underlined how an exposure time of less than 1 year was not enough to grasp a complex phenomenon as ICT effect on loneliness, but, at the same time, it is too long to measure small daily changes in older people's attitudes. The sample size was identified as a focal limitation in 3 experimental studies [33,34,36]: small samples (composed of 30 or 50 people) do not seem able to monitor pre-post differences in the relationships between ICT use and older people's emotional well-being, in particular loneliness. Also, Cotten et al [33] reported a similar limitation due to small sample size, despite the involvement of more than 205 individuals. Moreover, the issue of sample size should be linked to other limitations arising from the sampling process. In particular, the above mentioned authors [33,34,36] focused attention on the lack of detailed information about participant lifestyles, and, consequently, on the inability to perform integrative stratified analysis. Siegers et al [30] and Quinn [36] agreed with Cotten et al [33] about the lack of secondary analysis on participants' lifestyle context (eg, social environment, attitude toward active lifestyle, or participation in volunteering associations). Instead, from the point of view of Myhre et al [34] and Woodward et al [31], the main limitation of their studies concerned the adaptation of the randomization process due to the availability of the individuals to participate in the training course. Two studies [27,30] reported limitations due to the chosen measures, in particular regarding the use of self-reported scales to assess loneliness. Last, some bias was observed due to the translation of research and intervention tools from the original language to participants' language [29,32]. Some limitations coming from the structure of the interventions were identified by Fokkema and Knipscheer [28] and Morton et al [37]. In particular, the increase of social contacts between participants in the training course and their interactions with trainers and tutors could influence the effects on loneliness, although the study designs did not allow us to disentangle the effects of training from the trainer's visits. In the end, Morton et al [37] placed the emphasis on how the issue of the impact of online communication in older age and on the benefits coming from specialized courses would be transcended, since new generations of older people will already be ICT users.

Discussion

Principal Findings

The analysis of the findings of the reviewed studies highlights the growing interest in assessing the impact of SNSs on the social well-being of elderly people. The existence of a positive effect of ICT use on older people's loneliness seems to be confirmed, even if it is often a weak or not statistically significant effect. Moreover, the results underline how the use of ICTs has positive effects on the individual sense of self-worth, strength of personal identity, and self-esteem. The literature emphasizes the connection between these aspects of mental and psychological well-being and loneliness [42]. These results underline, once more, the hard work needed to analyze by experimental studies complex phenomena such as loneliness. The small number of experimental studies available in the literature on the effect of ICTs on social well-being and loneliness of older people and the limitations declared by the authors of the analyzed studies confirm this assumption. Indeed, many of these limitations concern the lack of control on the secondary variables that potentially influence the outcome, such as the effect of social contact during the training course, which is not easy to isolate [43]. In addition, the characteristics of lifestyle and level of social engagement of participants take on a specific relevance in the study designs. It is not by chance that the biggest limitation detected is small sample size, which does not ensure assessment of many of these secondary variables and, thus, a switch from specific findings to general assumptions on causal relationship. To confront the difficulty in defining the protocol, this review underlines how different methods are implemented to balance methodological accuracy and improvement of knowledge of the considered phenomenon. In addition, protocol designs were rarely supported by power analysis, pilot studies, or short qualitative studies [44].

Despite the low level of attrition in the selected studies (always below 10%), long- and medium-term studies involving older people need to take it into account carefully, because it could occur nonrandomly, thus posing challenges in the accuracy of results [45]. Furthermore, the heterogeneity of the selected population meets the debate on the definition of older people, conventionally defined as persons age 60 years. This definition is currently discussed in the literature [46]. Demographic trends and the literature underline how nowadays older age is setting in later [47]. This trend is confirmed by the sample composition: older people involved in the selected studies are generally aged 70 years or more.

In almost all of the reviewed studies, participants must have been early users of ICTs, but in some of them the availability of a PC is a fundamental participation criterion. This choice, although it alleviates the cost and management of the study, can

represent a source of bias in the sampling process. Indeed, in almost all cases the ICT use is self-reported by participants and may correspond to a different level of actual knowledge and ability in using the PC.

This review stresses the strong impact of the chosen recruitment process on the target population and sampling design. In particular, recruitment by flyers mostly involves people living at home and with a high level of autonomy, while recruitment in residential homes could find less healthy people with some functional limitations (eg, hearing or walking) [48]. These conditions could influence the social life of participants significantly. The assessment of loneliness was often not considered as a prerequisite to be involved in the study, even if it was the main outcome of several studies. Considering that loneliness is a subjective feeling, its detection at baseline allows a pre-post measurement of individual change and a comparison between individuals with different starting conditions. The sample must ensure the availability of this measurement, but often the reviewed studies underlined a bias in the sampling process: people who volunteered to participate in the study were often active people with fewer loneliness experiences. The widespread inclusion of at least one control group supports the analysis of the causal relationship between dependent (health and well-being of older people) and independent (ICT use) variables compared with daily life activities (passive control group) or a different socialization intervention (active control group). The detected trend of single follow-up underlines how these studies are often aimed at evaluating the short-term effect of an intervention. To capture long-term effects, some studies proposed further follow-ups after 6 months or 12 months or, as done by Fokkema and Knipscheer [28], after 3 years. The choice of Fokkema and Knipscheer [28] seems to be successful because it detected a changing trend in significant loneliness, underlining how loneliness trends are better detected by long-term rather than short-time evaluation. However, this choice needs a longer intervention characterized by additional tutoring activities between follow-ups to support ICT use by participants. Otherwise, as declared by Fokkema and Knipscheer [28] and Morton et al [37], a long-term intervention could facilitate the interfering effect of internal social activities on the self-perception of loneliness.

In the last 5 years, in line with the broader use of SNSs, the number of studies that have observed the socializing role of ICTs has grown. Loneliness is the primary measured outcome of individual social well-being and is often considered strongly related to other well-being aspects, such as personal life satisfaction, cognitive health conditions, and the ability to manage daily life events. The heterogeneity and number of measurement tools used confirm the theoretical and methodological debate on how complex phenomena (such as loneliness or social isolation) should be measured [49]. The variety of validated scales ensures the appropriateness of measure for quantitative studies, but it does not allow us to read the inside aspects of a personal feeling of loneliness. The choice to use multiple scales, done by many authors, pushes us to a deeper understanding of the issue [50]. Moreover, the use of validating scales ensures reliability of measurements, but each scale measures a specific aspect of loneliness phenomenon, and the variety of scales between studies reduces the comparability of results.

Limitations

Some limitations need to be considered in this review. First, the small amount of experimental articles in this area and their heterogeneity make it hard to produce a meta-analysis of the results and draw generalized conclusions. The decision to limit the literature search to articles published in English might have led to selection bias, although English is by far the leading language in this field of research. Second, the complex nature of the outcomes compared with the small size of the study samples limits the generalizability of the conclusions.

Conclusions

This review aimed to clarify how experimental studies improve the understanding of the causal relationships between older people's ICT use and their well-being concerning loneliness. Moreover, this review highlights the analysis of protocols applied to support the design of future research in this field. In particular, we compared 11 experimental and quasi-experimental studies published from 2002 to 2019. The characteristics of the analyzed study protocols (research questions, outcomes, evaluation tools, and treatments) highlighted difficulties in design, sampling, and management of the interventions. Nevertheless, despite the declared limitations, the overall findings are positive, highlighting the need for studying these issues with adequate methodological rigor. First, care is needed to discern the causal relationship between the dependent and independent variables from the effects of other intervening variables. Second, these difficulties affect the possibility of carrying out more than one follow-up—usually at the end of the intervention—and force the recruitment of small samples. The extensive use of the passive control group is an indirect effect of these difficulties. Moreover, this review underlines how the complex nature of loneliness and, even more, its relationship with ICT use, would require a complex and complete design of evidence-based studies, characterized by multivariable schemes and large sample sizes. On the other hand, randomized controlled studies allow the identification and analysis of causal relationships. The experimental studies included in the review show some difficulties and limitations in data collection due to the exclusive use of standardized tools to analyze the loneliness issue, in which individual feeling seems better detected by qualitative than quantitative methods.

The findings coming from the reviewed studies seem to confirm a beneficial effect—albeit weak—of ICT use on the well-being of older people in terms of reduced loneliness. The weakness of these results, along with the growing interest in the relationship between ICT use and loneliness in older age, draws attention to the need for development of further evidence-based studies. Future research in this field should take account of the need for studies with multidisciplinary design. The integration of clinical, psychological, and sociological research approaches would allow us to better verify primary and secondary outcomes of ICT use for older people's well-being, including loneliness. Moreover, quantitative protocol studies could benefit from a larger randomized sampling—better if supported by power

analysis—and by a short qualitative set of questions to improve the understanding and validity of the results.

Acknowledgments

This research was supported by Fondazione Cariplo under grant for “Bando 2017, ricerca scientifica: Ricerca sociale sull’invecchiamento: persone, luoghi e relazioni” (number 0946). This study was partially supported by Ricerca Corrente funding from the Italian Ministry of Health to Istituti di Ricovero e Cura a Carattere Scientifico—Istituto Nazionale di Ricovero e Cura degli Anziani.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Keywords and protocol characteristics.
[DOCX File , 17 KB - jmir_v23i3e23588_app1.docx]

Multimedia Appendix 2
Reference list of the international validated scales used in the reviewed studies.
[DOCX File , 22 KB - jmir_v23i3e23588_app2.docx]

References

1. World Population Ageing 2015. Population Division, Department of Economic and Social Affairs, United Nations. URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf [accessed 2021-01-30]
2. Holwerda TJ, Beekman ATF, Deeg DJH, Stek ML, van Tilburg TG, Visser PJ, et al. Increased risk of mortality associated with social isolation in older men: only when feeling lonely? Results from the Amsterdam Study of the Elderly (AMSTEL). Psychol Med 2012 Apr;42(4):843-853. [doi: 10.1017/S0033291711001772] [Medline: 21896239]
3. Smoyak SA. Loneliness: a sourcebook of current theory, research and therapy. J Psychosoc Nurs Ment Health Serv 1984 Jun 01;22(6):40-41. [doi: 10.3928/0279-3695-19840601-09]
4. Hansen T, Slagsvold B. Late-life loneliness in 11 European countries: results from the Generations and Gender Survey. Soc Indic Res 2015 Sep 29;129(1):445-464. [doi: 10.1007/s11205-015-1111-6]
5. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014 Sep;83(9):624-635 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
6. Zickuhr K, Madden M. Older adults and internet use. Washington: Pew Internet and American Life Project; 2012 Jun 06. URL: https://www.pewresearch.org/internet/2012/06/06/older-adults-and-internet-use/ [accessed 2021-01-30]
7. Michiels SI, Van Crowder L. Discovering the "Magic Box": local appropriation of information and communication technologies (ICTs). SD Dimens Sustain Dev Dep (SD), Food Agric Organ United Nations 2001:32-38 [FREE Full text]
8. Fernández-Ardèvol M, Rosales A. Older people, smartphones and WhatsApp. Smartphone Cult 2017:1. [doi: 10.4324/9781315307077-5]
9. Winstead V, Anderson WA, Yost EA, Cotten SR, Warr A, Berkowsky RW. You can teach an old dog new tricks: a qualitative analysis of how residents of senior living communities may use the web to overcome spatial and social barriers. J Appl Gerontol 2013 Aug;32(5):540-560 [FREE Full text] [doi: 10.1177/0733464811431824] [Medline: 25474761]
10. Khalaila R, Vitman-Schorr A. Internet use, social networks, loneliness, and quality of life among adults aged 50 and older: mediating and moderating effects. Qual Life Res 2018 Feb;27(2):479-489. [doi: 10.1007/s11136-017-1749-4] [Medline: 29210015]
11. Boyd D, Ellison N. Social network sites: definition, history, and scholarship. J Comput Commun 2007:1. [doi: 10.1111/j.1083-6101.2007.00393.x]
12. Nef T, Ganea RL, Müri RM, Mosimann UP. Social networking sites and older users: a systematic review. Int Psychogeriatr 2013 Jul;25(7):1041-1053. [doi: 10.1017/S1041610213000355] [Medline: 23552297]
13. Eggermont S, Vandebosch H, Steyaert S. Towards the desired future of the elderly and ICT: policy recommendations based on a dialogue with senior citizens. Poiesis Prax 2006 May 25;4(3):199-217. [doi: 10.1007/s10202-005-0017-9]
14. Gibson L, Moncur W, Forbes P, Arnott JA, Martin CG, Bhachu AS. Designing social networking sites for older adults. Proc 2010 Br Comput Soc Conf Human-Computer Interact 2010:1. [doi: 10.14236/ewic/hci2010.24]
15. Liu L, Cheung WK, Li X, Liao L. Aligning users across social networks using network embedding. Proc 25th Int Joint Conf Artificial Intel 2016:1.
16. Hutto C, Bell C, Farmer S, Fausset C, Harley L, Nguyen J, et al. Social media gerontology: understanding social media usage among older adults. Web Intell Agent Syst 2015 Apr 29;13(1):69-87. [doi: 10.3233/web-150310]

17. Schehl B, Leukel J, Sugumaran V. Understanding differentiated internet use in older adults: a study of informational, social, and instrumental online activities. Comput Human Behav 2019 Aug;97:222-230. [doi: 10.1016/j.chb.2019.03.031]
18. Gaia A, Sala E, Cerati G. Social networking sites use and life satisfaction. A quantitative study on older people living in Europe. Eur Soc 2020 May 19:1-21. [doi: 10.1080/14616696.2020.1762910]
19. Jackson M, Cox D. The principles of experimental design and their application in sociology. Ann Rev Sociol 2013 Jul 30;39(1):27-49. [doi: 10.1146/annurev-soc-071811-145443]
20. Walsh K, Scharf T, Keating N. Social exclusion of older persons: a scoping review and conceptual framework. Eur J Ageing 2016 Oct 11;14(1):81-98. [doi: 10.1007/s10433-016-0398-8]
21. Ige J, Gibbons L, Bray I, Gray S. Methods of identifying and recruiting older people at risk of social isolation and loneliness: a mixed methods review. BMC Med Res Methodol 2019 Aug 29;19(1):181 [FREE Full text] [doi: 10.1186/s12874-019-0825-6] [Medline: 31464586]
22. Fan Q. Utilizing ICT to prevent loneliness and social isolation of the elderly: a literature review. Cuad Trab Soc 2016:1. [doi: 10.5209/CUTS.251771]
23. Nervik T, Dahl Y, Kofod-petersen A. The value of role models in design of social media for elderly. 23rd Int Conf Eur Fed Med Inform. 2011. URL: https://folk.idi.ntnu.no/anderpe/publications/MIE-11-TN-YD-AKP.pdf [accessed 2021-01-30]
24. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
25. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
26. Gorman M. Development and the rights of older people. In: Randel J, German T, Ewing D, editors. The Ageing and Development Report: Poverty, Independence and the World's Older People. London: Taylor & Francis; 2017.
27. White H, McConnell E, Clipp E, Branch LG, Sloane R, Pieper C, et al. A randomized controlled trial of the psychosocial impact of providing internet training and access to older adults. Aging Ment Health 2002 Aug;6(3):213-221. [doi: 10.1080/13607860220142422] [Medline: 12217089]
28. Fokkema T, Knipscheer K. Escape loneliness by going digital: a quantitative and qualitative evaluation of a Dutch experiment in using ECT to overcome loneliness among older adults. Aging Ment Health 2007 Sep;11(5):496-504. [doi: 10.1080/13607860701366129] [Medline: 17882587]
29. Shapira N, Barak A, Gal I. Promoting older adults' well-being through Internet training and use. Aging Ment Health 2007 Sep;11(5):477-484. [doi: 10.1080/13607860601086546] [Medline: 17882585]
30. Slegers K, Van Boxtel MPJ, Jolles J. Effects of computer training and Internet usage on the well-being and quality of life of older adults: a randomized, controlled study. J Gerontol B Psychol Sci Soc Sci 2008 May;63(3):P176-P184. [Medline: 18559683]
31. Woodward AT, Freddolino PP, Blaschke-Thompson CM, Wishart DJ, Bakk L, Kobayashi R, et al. Technology and aging project: training outcomes and efficacy from a randomized field trial. Ageing Int 2010 Sep 28;36(1):46-65. [doi: 10.1007/s12126-010-9074-z]
32. Blažun H, Saranto K, Rissanen S. Impact of computer training courses on reduction of loneliness of older people in Finland and Slovenia. Comput Human Behav 2012 Jul;28(4):1202-1212. [doi: 10.1016/j.chb.2012.02.004]
33. Cotten SR, Ford G, Ford S, Hale TM. Internet use and depression among older adults. Comput Human Behav 2012 Mar;28(2):496-499. [doi: 10.1016/j.chb.2011.10.021]
34. Myhre JW, Mehl MR, Glisky EL. Cognitive benefits of online social networking for healthy older adults. J Gerontol Ser B Psychol Sci Soc Sci 2016 Mar 16:gbw025. [doi: 10.1093/geronb/gbw025]
35. Larsson E, Padyab M, Larsson-Lund M, Nilsson I. Effects of a social internet-based intervention programme for older adults: an explorative randomised crossover study. Br J Occup Ther 2016 Jul 19;79(10):629-636. [doi: 10.1177/0308022616641701]
36. Quinn K. Cognitive effects of social media use: a case of older adults. Soc Media Soc 2018 Jul 19;4(3):205630511878720. [doi: 10.1177/2056305118787203]
37. Morton TA, Wilson N, Haslam C, Birney M, Kingston R, McCloskey L. Activating and guiding the engagement of seniors with online social networking. J Aging Health 2016 Aug 01:898264316664440. [doi: 10.1177/0898264316664440] [Medline: 27530332]
38. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
39. Guest G, Namey E, Mitchell M. Collecting qualitative data: a field manual for applied research. Collect Qual Data A F Man Appl Res 2017:1. [doi: 10.4135/9781506374680]
40. Russell D, Cutrona CE, Rose J, Yurko K. Social and emotional loneliness: an examination of Weiss's typology of loneliness. J Pers Soc Psychol 1984 Jun;46(6):1313-1321. [doi: 10.1037//0022-3514.46.6.1313] [Medline: 6737214]
41. Larsson J, Björkdahl A, Esbjörnsson E, Sunnerhagen KS. Factors affecting participation after traumatic brain injury. J Rehabil Med 2013 Sep;45(8):765-770 [FREE Full text] [doi: 10.2340/16501977-1184] [Medline: 24002312]

42. Kim J, Lee S, Chun S, Han A, Heo J. The effects of leisure-time physical activity for optimism, life satisfaction, psychological well-being, and positive affect among older adults with loneliness. Ann Leis Res 2016 Oct 03;20(4):406-415. [doi: 10.1080/11745398.2016.1238308]
43. Gouveia O, Matos A, Schouten M. Social networks and quality of life of elderly persons: a review and critical analysis of literature. Rev Bras Geriatr e Gerontol 2016;19(6):1030-1040. [doi: 10.1590/1981-22562016019.160017]
44. Lüders M, Brandtzæg PB. "My children tell me it's so simple": a mixed-methods approach to understand older non-users' perceptions of Social Networking Sites. New Media Soc 2016 Jul 09;19(2):181-198. [doi: 10.1177/1461444814554064]
45. Mody L, Miller D, McGloin J, Freeman M, Marcantonio E, Magaziner J, et al. Recruitment and retention of older adults in aging research. J Am Geriatr Soc 2008 Dec;56(12):2340-2348 [FREE Full text] [doi: 10.1111/j.1532-5415.2008.02015.x] [Medline: 19093934]
46. Gobbens RJ, Luijkx KG, Wijnen-Sponselee MT, Schols JM. Toward a conceptual definition of frail community dwelling older people. Nurs Outlook 2010;58(2):76-86. [doi: 10.1016/j.outlook.2009.09.005] [Medline: 20362776]
47. Arnoldussen IAC, Sundh V, Bäckman K, Kern S, Östling S, Blennow K, et al. A 10-year follow-up of adiposity and dementia in Swedish adults aged 70 years and older. J Alzheimers Dis 2018;63(4):1325-1335. [doi: 10.3233/JAD-180099] [Medline: 29758945]
48. Falconer M, O'Neill D. Profiling disability within nursing homes: a census-based approach. Age Ageing 2007 Jan 25;36(2):209-213. [doi: 10.1093/ageing/afl185]
49. Zavaleta D, Samuel K, Mills CT. Measures of social isolation. Soc Indic Res 2016 Feb 2;131(1):367-391. [doi: 10.1007/s11205-016-1252-2]
50. Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. Ann Intern Med 2000 Sep 19;133(6):455-463 [FREE Full text] [doi: 10.7326/0003-4819-133-6-200009190-00014] [Medline: 10975964]

Abbreviations

ICT: information and communication technology
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SNS: social networking site
UCLA: University of California Los Angeles

Edited by G Eysenbach; submitted 18.08.20; peer-reviewed by M Razai, E Toki; comments to author 22.09.20; revised version received 05.10.20; accepted 12.01.21; published 01.03.21.

Please cite as:
Casanova G, Zaccaria D, Rolandi E, Guaita A
The Effect of Information and Communication Technology and Social Networking Site Use on Older People's Well-Being in Relation to Loneliness: Review of Experimental Studies
J Med Internet Res 2021;23(3):e23588
URL: https://www.jmir.org/2021/3/e23588
doi:10.2196/23588
PMID:33439127

Review

Compliance With Mobile Ecological Momentary Assessment of Self-Reported Health-Related Behaviors and Psychological Constructs in Adults: Systematic Review and Meta-analysis

Marie T Williams[1*], BAppSc, PhD; Hayley Lewthwaite[1,2*], BAppSc, BHlthScHons, PhD; François Fraysse[3], PhD; Alexandra Gajewska[3], BClinExPhys, BAHons, Dip Ed; Jordan Ignatavicius[3], BPhysio; Katia Ferrar[3*], BAppSc, BHlthScHons, PhD

[1]Innovation, Implementation And Clinical Translation in Health, Allied Health and Human Performance, University of South Australia, Adelaide, Australia

[2]Department of Kinesiology and Physical Education, Faculty of Education, McGill University, Montreal, QC, Canada

[3]Alliance for Research in Exercise, Nutrition and Activity, Allied Health and Human Performance, University of South Australia, Adelaide, Australia

[*]these authors contributed equally

Corresponding Author:
Marie T Williams, BAppSc, PhD
Innovation, Implementation And Clinical Translation in Health
Allied Health and Human Performance
University of South Australia
City East Campus, North Terrace
Adelaide, 5000
Australia
Phone: 61 8 8302 1153
Email: Marie.Williams@unisa.edu.au

Abstract

Background: Mobile ecological momentary assessment (mEMA) permits real-time capture of self-reported participant behaviors and perceptual experiences. Reporting of mEMA protocols and compliance has been identified as problematic within systematic reviews of children, youth, and specific clinical populations of adults.

Objective: This study aimed to describe the use of mEMA for self-reported behaviors and psychological constructs, mEMA protocol and compliance reporting, and associations between key components of mEMA protocols and compliance in studies of nonclinical and clinical samples of adults.

Methods: In total, 9 electronic databases were searched (2006-2016) for observational studies reporting compliance to mEMA for health-related data from adults (>18 years) in nonclinical and clinical settings. Screening and data extraction were undertaken by independent reviewers, with discrepancies resolved by consensus. Narrative synthesis described participants, mEMA target, protocol, and compliance. Random effects meta-analysis explored factors associated with cohort compliance (monitoring duration, daily prompt frequency or schedule, device type, training, incentives, and burden score). Random effects analysis of variance ($P \leq .05$) assessed differences between nonclinical and clinical data sets.

Results: Of the 168 eligible studies, 97/105 (57.7%) reported compliance in unique data sets (nonclinical=64/105 [61%], clinical=41/105 [39%]). The most common self-reported mEMA target was affect (primary target: 31/105, 29.5% data sets; secondary target: 50/105, 47.6% data sets). The median duration of the mEMA protocol was 7 days (nonclinical=7, clinical=12). Most protocols used a single time-based (random or interval) prompt type (69/105, 65.7%); median prompt frequency was 5 per day. The median number of items per prompt was similar for nonclinical (8) and clinical data sets (10). More than half of the data sets reported mEMA training (84/105, 80%) and provision of participant incentives (66/105, 62.9%). Less than half of the data sets reported number of prompts delivered (22/105, 21%), answered (43/105, 41%), criterion for valid mEMA data (37/105, 35.2%), or response latency (38/105, 36.2%). Meta-analysis (nonclinical=41, clinical=27) estimated an overall compliance of 81.9% (95% CI 79.1-84.4), with no significant difference between nonclinical and clinical data sets or estimates before or after data exclusions. Compliance was associated with prompts per day and items per prompt for nonclinical data sets. Although widespread heterogeneity existed across analysis (I^2>90%), no compelling relationship was identified between key features of mEMA protocols representing burden and mEMA compliance.

Conclusions: In this 10-year sample of studies using the mEMA of self-reported health-related behaviors and psychological constructs in adult nonclinical and clinical populations, mEMA was applied across contexts and health conditions and to collect a range of health-related data. There was inconsistent reporting of compliance and key features within protocols, which limited the ability to confidently identify components of mEMA schedules likely to have a specific impact on compliance.

(J Med Internet Res 2021;23(3):e17023) doi:10.2196/17023

KEYWORDS

mobile momentary ecological assessment; adult; compliance; systematic review; meta-analysis; mobile phone

Introduction

Background

Ecological momentary assessment (EMA) is a survey method that allows collection of data on participant behaviors, affect, and perceptual experiences in real-time (momentary) and real-life environments (ecological) [1]. In its original form, EMA required pen and paper diaries or logs to be completed on random (signal) or fixed (interval) time-based schedules or in response to a specific target behavior, psychological or social event (event-based). With the advent of handheld technologies, mobile EMA (mEMA) and increasingly mobile ecological momentary interventions (mEMIs) can be completed through automated schedules via handheld devices such as tablets and mobile phones.

As mEMA or mEMI have the potential to capture data in real time, the level of recall bias is potentially reduced. In addition, contextual (where and who the respondent is with) and antecedents to the specific target behavior or psychological construct can be obtained [1,2]. As a survey approach, mEMA or mEMI has undeniable utility, but data are dependent on participants consistently responding to the mEMA or mEMI schedule (compliance) [3]. Although electronically delivered surveys to personal mobile devices provide a means of time or date stamping and limit the possibility of hoarding, back and forward filling [4], concerns have been raised about protocol burden, missing data (especially if systematic), mindless answering, and survey habituation when lengthier questionnaires can be circumvented by a no response to initial questions [2]. EMA data with low compliance rates are unlikely to be ecologically valid; however, it is also possible to have good individual compliance with data of questionable accuracy [5,6].

In the last 5 years, there have been at least 10 systematic reviews focused on EMA and/or reporting aspects of compliance to EMA schedules in youth (<18 years [7-9]; <22 years [10]), mixed youth and adult cohorts [11-13], or specific adult populations [5,14-16]. Compliance with EMA in youth (nonclinical and clinical samples) has been reported to range between 44% and 96% [8-10] and in mixed youth and adult cohorts, between 23% and 94% [11-14]. Reports of compliance in specific adult clinical populations range from 21% to 99% (chronic pain, 21%-99% [15]; psychotic disorders, 78%-86% [16]; substance use, 75%, (95% CI 72.37-77.65) [5].

Although Stone and Shiffman [17] have highlighted the need for explicit reporting of compliance in their original reporting guidelines for EMA, recurring issues relating to the reporting of compliance include (1) missing, incomplete, or ambiguous data; (2) heterogeneity in reporting; (3) impact of data exclusions; and (4) combining traditional (paper-based) and mEMA data [5]. Participant compliance with mEMA or mEMI—in theory—is related to the total protocol burden, which is a function of monitoring duration, frequency and complexity of prompts, and familiarity with the technology. However, as Jones et al [5] note, to date, there is little compelling, systematic evidence to support an association between EMA burden and compliance rates. These issues make it difficult to determine which, if any, features of EMA protocols positively or negatively influence compliance to EMA schedules.

The purpose of this systematic review is to guide the development of an mEMA protocol, which could be used for future studies of health-related behaviors and psychological constructs (including symptoms) in adults with and without chronic disease. The primary question for this systematic review is as follows: In adult nonclinical and clinical populations, which factors are associated with increased compliance to mEMA protocols for collection of health-related behaviors and psychological constructs (including symptoms)?

Objectives

The objectives of this systematic review were to describe:

1. Health-related behaviors and psychological constructs assessed using mEMA
2. mEMA protocol and compliance reporting
3. Associations between key components of mEMA protocols and participant compliance

Methods

Search Registration

The search strategy and review protocol were registered prospectively with the International Prospective Register of Systematic Reviews (PROSPERO 2016: CRD42016051726).

Eligibility

Observational studies (cohort, cross-sectional) of mEMA in adults (>18 years of age) were eligible for inclusion in this review if these (1) reported participant compliance with mEMA; (2) were a primary study published in English between 2006 and 2016 inclusive; (3) included adults (≥18 years) either apparently healthy (nonclinical population) or with health conditions (clinical population); and (4) collected mEMA data using mobile devices as a primary or secondary outcome. References were excluded if these were (1) experimental designs investigating intervention efficacy; (2) duplicate publications or secondary analysis of the same data set; or (3) conference

abstracts, protocols, commentaries (editorials or letters), or systematic or narrative reviews.

Information Sources and Search Strategy

A range of electronic databases were searched to identify eligible studies: AMED (Allied and Complementary Medicine), CINAHL, Cochrane Library and CENTRAL (Cochrane Central Register of Controlled Trials), Embase, MEDLINE (including epub ahead of print), PsycINFO, Scopus, and Web of Science. An academic librarian (Carole Gibbs, University of South Australia) assisted with the development of the search strategy regarding conceptualization, operators (operational terms), and limiters [18] with the final search undertaken during a single week. Search terms and associated MeSH (Medical Subject Heading) alternatives, which were adapted for use in all databases, related to the population (adults), assessment (mEMA), and outcomes of interest (health behaviors, perceptual experiences including symptoms, affect or mood). Key search terms included "ecological momentary assessment," "EMA," "mobile ecological momentary assessment," "mEMA," "electronic diary," "SMS or short message service," "prompting," "text messaging," "health behaviour," "symptom," and "adult." Reference lists of included studies and systematic reviews identified during the search were reviewed to identify additional potentially relevant studies.

Study Selection

The titles and abstracts of studies identified from the search process were screened against a priori eligibility criteria and full-text versions imported into Covidence (Covidence systematic review software, Veritas Health Innovation). Both screening steps were undertaken by individual members of the research team working in pairs (AG and MW, HL and FF) with each person completing the task independently, before meeting with their partner to compare results and resolve disagreements (consensus).

Data Collection

A data extraction template was prospectively developed; it was guided by the Checklist for Reporting EMA studies proposed by Liao et al [10] and pilot-tested on 5 randomly selected eligible studies. Working in pairs (AG and MW, JI and KF, HL and FF), individual members of the research team extracted all data before meeting with their partner to compare results and resolve disagreements by discussion. As this review aims to describe the features of mEMA schedules associated with increased mEMA protocol adherence, assessment of methodological bias was not planned.

Data Items

Data were extracted across 4 domains:

Publication demographics: title, authors, year of publication.

Participants: recruitment source, medical condition or diagnosis (clinical populations), sample size (enrolled, attrition or withdrawn and included in analysis), and age (mean/median, SD).

mEMA protocol: target behavior or psychological construct, mobile device type (PDA, palmtop computer, electronic diary, mobile or smartphone, tablet, other), participant training (yes/no), provision of incentives (course credit, financial, other, or none), incentive thresholds (yes/no) monitoring duration (days), prompt type (random signal, interval, event-based), frequency per day, number of questions/items per prompt type (reported or estimated from information reported in studies), strategy to deal with unanswered prompts, and time allowed for survey response. Where authors did not report the number of items per prompt type, but rather included descriptions of standardized instruments which were converted to mEMA survey items, a full version of the standardized instrument was accessed, and number of items calculated.

mEMA compliance: verbatim (or where possible calculated from reported data), participant completion (number included in analysis, data exclusions), criteria/thresholds for mEMA data, number of prompts delivered/answered per person/cohort (planned, actual, average, range), and response latency as time (mean, SD) [8,10].

Data Management

Data were tabulated to provide descriptive summaries. The mEMA surveys commonly included multiple questions reflecting behavioral or psychological constructs. Although the authors of mEMA studies did not always specify the primary outcome for these observational studies, most studies explicitly reported the key variable of interest for mEMA, which we interpreted to be the primary mEMA target. Where other data were also collected by the same mEMA survey, we denoted those as secondary mEMA targets. The primary mEMA target of studies was identified, and studies were grouped and reported according to two broad domains: (1) behavior (eg, dietary, physical activity, and smoking) and (2) psychological construct (eg, affect, cognition, and sensations/symptoms). For each domain, a narrative synthesis was used to summarize participants, mEMA protocol, and compliance data for nonclinical and clinical data sets.

With the exception of device type, where possible, we adopted the operationalization of variables common to Wen et al [9] or Jones et al [5] unless the distribution of our data resulted in very unbalanced cells or our data could provide greater resolution. Potential mEMA protocol factors related to compliance were categorized for analysis. *Monitoring duration* was categorized as follows: <7 days, >7 days to <14 days, or >14 days. *Prompt frequency* was grouped as follows: 1-3 prompts per day; 4-5 prompts per day; or ≥6 prompts per day. Minimum *items per prompt* were categorized as follows: ≤5, >5 to ≤9.5, >9.5 to ≤26, and >26. *Device type* was categorized as mobile phone, PalmPilot/PDA, or other. The reporting of *training or familiarization sessions or provision of incentives* were dichotomized as yes/no or labeled as not reported.

Given ongoing concerns about the burden imposed by EMA schedules and compliance, in addition to these individual factors, we explored a novel composite metric to reflect aspects previously identified as possible contributing factors (monitoring duration, frequency, type, and complexity of prompts).

Where possible, a mEMA *burden score* was calculated for each study by multiplying:

- the total monitoring duration in days (d; all days included in all waves)
- by the maximum frequency of time-based prompts (random and interval) per day (f)
- by the minimum number of compulsory questions/items within all prompts per day (i) and
- by a weighting reflecting the number of prompt types scheduled per day (w; eg, time-based [signal or interval] and/or event-based) with each prompt type weighted as 1 (min weight=1, max=3).

For example, the mEMA burden score for a 14-day monitoring schedule *(d)*, where 5 random signal prompts were delivered per day (*f*), with each prompt requiring responses to a minimum of 12 items/questions (*I*; 60 items in total per day), would be 840. If event-based prompts (irrespective of the number of items within the prompt) were added to this schedule (*w*), the burden score would rise to 1680. *Burden scores* were calculated and reported in quartiles: 0 to 283.5, 284 to 810, 811 to 1806, or ≥1807.

Meta-analysis

Random effects restricted maximum likelihood estimator meta-analyses were undertaken using the approach reported by Jones et al [5] and Wen et al [9], with both authors advising to assist in accurate replication. All statistical analyses were conducted using JASP (Jeffreys's Amazing Statistics Program, version 0.9.2; 2019). Studies were included in the meta-analysis if they reported all data necessary for the meta-analysis procedure and cohort compliance (%) could be extracted before data exclusions when possible. Sensitivity analysis was conducted to explore the impact of compliance rates reported before and after data exclusion. The effect sizes (ESs) were calculated by logit transforming the proportion of completed prompts (ie, compliance rates; proportion/[1−proportion]). SEs were then estimated using the following equation:

$$\sqrt{([1/np]+[1/n\{1-p\}])}$$

Where, *n* is the sample size and *p* is the proportion.

To adjust for clustering within participants, the SE was adjusted by the effective sample size (ESS). The ESS equation is as follows:

$$kn/(1+[k-1]\ ICC)$$

Where, k is the number of study prompts, n is the participant number, ICC is either the reported intraclass correlation coefficient (ICC) or the SD of reported compliance, and p is the proportion of completed prompts.

For studies that did not report SD data, sensitivity analyses were conducted by computing the SEs using the 25 and 75 percentiles of available SDs. The sensitivity analyses did not show any differences. Therefore, analysis used imputed median SD (where the original SD was not reported). To aid interpretation, inverse logit transformation was conducted to enable reporting of proportions. The I^2 statistic was used to quantify heterogeneity across the ES. Pooled compliance rates were initially explored for combined nonclinical and clinical data sets and then compared between nonclinical and clinical studies.

To explore the relationships between the pooled compliance rates (nonclinical and clinical data sets) and EMA protocol factors (ie, monitoring duration, prompt frequency, device type, training, incentives, and burden score), random effects analysis of variance was conducted as part of the meta-analysis program. Moderator analyses were conducted separately for nonclinical and clinical pooled compliance.

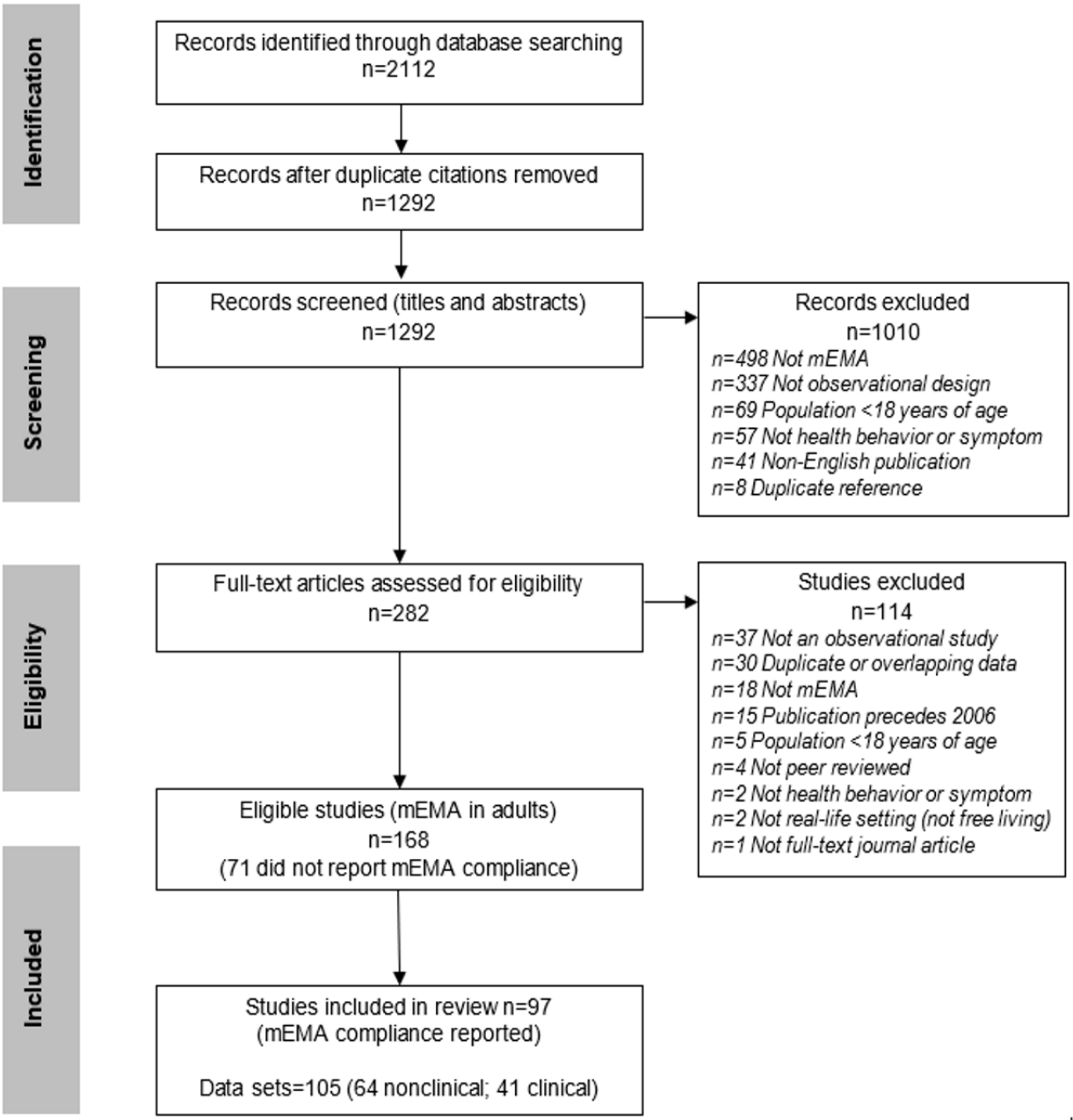
Results

Overview

Figure 1 presents the outcome of the search strategy. Of the 282 studies reviewed as full text, 168/282 (59.6%) included mEMA; however, 42.3% (71/168) were excluded because mEMA compliance was not reported. The majority of the 97 studies retained for this review comprised studies that recruited or reported a single nonclinical group (61/97, 63%) or a clinical (31/97, 32%) group. Two studies included 2 [19] or 3 clinical groups [20]. In addition, 3 studies included clinical and nonclinical comparator groups (4 groups [21], 2 groups [22,23]). Overall, 105 data sets were included in this review (nonclinical: 64/105, 61%; clinical: 41/105, 39%). A description of all included data sets is presented in Multimedia Appendix 1 [19-114].

A total of 44,796 participants were included in the analyses (nonclinical: 42,338/44,796, 94.51%; clinical: 2431/44,796, 5.43%) with a median sample size of 62 (nonclinical: n=89; clinical: n=40; Multimedia Appendix 2). Two data sets (nonclinical) were outliers because of the sample size (n=21,947; n=11,572) [24,25]. The main sources of recruitment for nonclinical data sets were educational institutions (30/64, 47%) and community (26/64, 41%), whereas clinical data sets were predominantly recruited from medical/health services (21/41, 51%) and community (17/41, 41%). For clinical data sets, the most common health conditions were psychiatric or mental health (12/41, 29%), chronic pain and fibromyalgia (6/41, 15%), and eating disorders (5/41, 12%). Multimedia Appendix 2 presents a summary of the study characteristics grouped by primary mEMA target.

Figure 1. Search strategy process and final outcomes (hand searching of reference list-eligible studies and review papers did not identify additional studies to those returned by database searches). mEMA: mobile ecological momentary assessment.



Objective 1: Health-Related Behaviors and Psychological Constructs Assessed With mEMA

Using the primary mEMA target, data sets were grouped into 2 broad domains: *Behavior* or *Psychological construct*. Within the *Behavior* domain, the *Other* category reflects single studies (7), where the primary mEMA target did not align with more common behavior targets (social interactions/activities [26,27], sexual [28], leisure [29], nonsuicidal self-injurious [30], HIV prevention [31], and oral behaviors) [32].

The most frequent primary mEMA target across all domains for nonclinical and clinical data sets was affect (31/105, 29.5% of data sets; nonclinical n=15/64, 14%, clinical n=16/41,15%). The most common primary mEMA target in nonclinical data sets (n=64) reflected the Behavior domain (total 38/64, 59%), whereas clinical data sets (n=41) reflected the Psychological domain (total 32/41, 78%).

With the exception of 1 clinical study (fatigue) [33], the remaining data sets included mEMA items/questions beyond the primary mEMA target. The most frequent secondary targets assessed were affect (50/105, 47.6%), social environment (33/105, 31.4%), physical activity (25/105, 23.8%), cognition (24/105, 22.8%), and physical environment (20/105, 19%). Multimedia Appendix 2 presents a summary of secondary mEMA targets and participant characteristics grouped by the primary mEMA target.

Objective 2: mEMA Protocol and Compliance Reporting

Multimedia Appendix 3 presents a summary of mEMA protocols grouped by primary mEMA target. Among the included studies,

mEMA data were most commonly collected using handheld computer/PDAs (61/105, 58.1%) with mobile phones accounting for approximately one-third (37/105, 35.2%). Participant training in mEMA was reported by most studies (nonclinical: 49/64, 77%; clinical: 35/41, 85%). The provision of incentive (financial or other) was more frequent in nonclinical protocols (nonclinical: 46/64, 72%; clinical: 20/41, 49%).

Across all data sets (n=105), the median monitoring duration for mEMA protocols was 7 days (range: 1-182 days), with durations differing between nonclinical (median 7 days, range 1-49 days) and clinical protocols (median 12 days, range 1-182 days). Most studies included a single prompt type (overall data sets: 69/105, 65.7%; nonclinical: 40/64, 63%; clinical: 29/41, 71%), with random signals being the most common in nonclinical protocols (49/64, 77%) and interval in clinical protocols (25/41, 61%). Of the remaining study protocols, 23% (24/105) of studies included 2 prompt types and 11% (12/105) protocols included all 3 prompt types (random signal, interval, and event-based). The frequency of time-based prompts (signal or interval) ranged from 1 to 42 per day (median: nonclinical=5, range 1-36; clinical=4, range=1-42). The number of specific questions/items within a standard prompt varied markedly across study protocols; it ranged between 1 and 73 (median: nonclinical=10; clinical=8).

Table 1 presents a summary of reporting for compliance metrics for mEMA time-based prompts (ie, signal and fixed prompts). Participant attrition (dropout) rates were reported or could be calculated for half of the 105 data sets (nonclinical: 31/64, 48%; clinical: 22/41, 54%). Less than half of the data sets reported the number of prompts delivered (overall: 22/105, 21%; nonclinical: 14/64, 22%; clinical: 8/41, 20%) or answered (overall: 43/105, 41%; nonclinical: 29/64, 45%; clinical: 14/41, 34%). Approximately one-third of the data sets reported a criterion for valid mEMA data or reasons for data exclusions (overall: 37/105, 35%; nonclinical: 25/64, 39%; clinical: 12/41, 29%). Criteria for valid EMA data fell into 2 main groups, with the most common based on assessment completion (ie, specified threshold for number of prompts completed per day or percentage of overall compliance), followed by response latency period threshold (eg, prompt required to be answered within 30 min). Of the data sets reporting a criterion for response time (overall: 38/105, 36%; nonclinical: 16/64, 25%; clinical: 22/41, 54%), this ranged from 1.5 to 60 min (median 15 min; Multimedia Appendix 3). Other reasons for data exclusion were based on specific time of day prompts (excluding the first or last of the day), technical malfunctions, or unspecified (eg, general statements on participants' limited or poor compliance).

Of the 105 data sets, 82/105 (78.1%) reported compliance using a single metric (cohort, average per person or other), with compliance at the cohort level most common (overall: 62/105, 59%; nonclinical: 34/64, 53%; clinical: 28/41, 68%). Compliance was less frequently reported using the single metric of average per person (overall: 20/105, 19%; nonclinical: 14/64, 22%; clinical: 6/41, 15%) or compliance for both cohort and average per person (overall: 18/105, 17%; nonclinical: 12/64, 19%; clinical: 6/41, 15%). The remaining data sets (n=5; nonclinical: n=4, clinical: n=1) reported compliance after combining event/time-based signals [34] or separate tasks [35], number of completed protocol days [36], total number of prompts (data) available [37], or proportion of completed questions/items per prompt [38]. Cohort compliance reported before data exclusions ranged from 38% to 98% (median 82%) and after data exclusions from 50% to 97% (median 81%; Table 1).

Table 1. Summary of mobile ecological momentary assessment (mEMA) compliance reporting.

Primary mEMA[a] target	NC[b] or C[c] (n)	Reported N=data sets (%)							Cohort compliance (%)	
		Attrition rate	Total prompts delivered	Total prompts answered	Criteria for valid data	Compliance predata exclusions	Compliance postdata exclusions	Average per-person compliance	Predata exclusion, median (range)	Postdata exclusion, median (range)
Smoking	NC (12)	8 (66)	4 (33)	5 (42)	4 (33)	7 (58)	1 (8)	5 (42)	83 (69-93)	83 (74-91)
	C (1)	1 (100)	1 (100)	1 (100)	0 (0)	1 (100)	0 (0)	1 (100)	68 (NA[d])	N/A[e]
Alcohol	NC (8)	3 (37)	1 (12)	3 (37)	3 (37)	4 (50)	4 (50)	2 (25)	90 (86-97)	79 (69-80)
	C (0)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Eating behaviors	NC (10)	6 (60)	2 (20)	5 (50)	5 (50)	4 (40)	3 (3)	5 (50)	90 (40-96)	67 (50-71)
	C (3)	2 (66)	1 (33)	2 (66)	1 (33)	2 (66)	1 (33)	0 (0)	N/A	78 (68-87)
Physical activity	NC (5)	1 (20)	1 (20)	4 (80)	0 (0)	3 (60)	0 (0)	3 (60)	82 (75-95)	N/A
	C (1)	0 (0)	0 (0)	0 (0)	0(0)	1 (100)	0 (0)	0 (0)	N/A	97 (NA)
Other	NC (3)	0 (0)	2 (66)	3 (100)	0 (0)	2 (66)	0 (0)	2 (66)	61 (38-84)	N/A
	C (4)	4 (100)	0 (0)	3 (75)	1 (25)	2 (50)	1 (25)	2 (50)	74 (72-74)	N/A
Personality traits	NC (7)	4 (57)	1 (14)	3 (42)	3 (42)	3 (42)	1 (14)	3 (42)	75 (55-90)	N/A
	C (0)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Affect	NC (15)	7 (46)	2 (13)	5 (33)	9 (60)	6 (40)	6 (40)	7 (46)	78 (63-90)	77 (73-81)
	C (16)	9 (56)	2 (12)	5 (31)	7 (44)	5 (31)	6 (37)	9 (56)	80 (69-96)	83 (79-87)
Cognitions	NC (2)	1 (50)	0 (0)	1 (50)	0 (0)	2 (100)	0 (0)	0 (0)	83 (77-89)	N/A
	C (0)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Symptoms	NC (2)	1 (50)	1 (50)	0 (0)	1 (50)	0 (0)	1 (50)	1 (50)	N/A	N/A
	C (16)	6 (37)	4 (25)	3 (19)	3 (19)	11 (69)	4 (25)	4 (25)	90 (68-98)	86 (86-93)
Total	NC (64)	31 (48)	14 (22)	29 (45)	25 (39)	31 (48)	16 (25)	28 (44)	82 (38-97)	74 (50-91)
	C (41)	22 (54)	8 (20)	14 (34)	12 (29)	22 (54)	12 (29)	16 (39)	80 (68-98)	87 (68-97)
	t (105)	53	22	43	37	53	28	44	82 (38-98)	81 (50-97)
	%	50.4	20.9	40.9	35.2	50.4	26.6	41.9	N/A	N/A

[a]mEMA: mobile ecological momentary assessment.

[b]NC: nonclinical.

[c]C: clinical.

[d]NA: not available as domain includes a single study.

[e]N/A: not applicable.

Question 3: Associations Between Key Features of mEMA Protocols and mEMA Compliance

Of the 105 data sets included in this review, 65% reported sufficient data for inclusion in the meta-analysis (n=68 data sets: 41/105 [39%] ES nonclinical and 27/105 [26%] ES clinical); Multimedia Appendix 1) [20,21,23,26,27,29-31,33,36,39-90]. The remaining data sets did not report cohort compliance but reported average per-person compliance [19,24,25,91-106,28,32] or other [34,35,37,38,107-110], or where cohort compliance was reported, a variable required for the meta-analysis was not [111-114].

The overall compliance rate across all 68 ESs was 81.9% (95% CI 79.1-84.4). There was sizable heterogeneity across the compliance rates (I^2=98). Sensitivity analysis exploring the impact of pre and postdata exclusion compliance rates showed no significant difference (*P*=.67*;* before exclusion: n=50, 81.6%; after exclusion: n=18, 82.8%). There was no significant difference (*P*=.16) between the pooled compliance of nonclinical studies (80.4%; 95% CI 76.1-83.9; I^2=98.6) and clinical studies (84.2%; 95% CI 80.1-87.4; I^2=95.7). Three studies included more than 1 data set and reported compliance ESs for each (data sets n=2 [23], n=3 [20], and n=4 [21]). Sensitivity analysis was undertaken to explore the impact of double counting of mEMA protocol factors within the meta-analysis, where multiple ESs were reported within single studies. When a single ES was retained for each of these studies (lowest ES of the 2 [23], median of 3 [20], ES closest to the average for 4 [21]), the

pooled 62 ESs (81.3%, 95% CI 78.2-84.2) and reported variance (I^2=98) were essentially the same as the full data set (68 ESs: 81.9%; 95% CI 79.1-84.4; I^2=98). To ensure that subgroup analysis was not affected, all analyses were conducted without duplicate ESs, and all relationships were consistent with those of the full data set.

For nonclinical studies, 2 factors (prompt frequency and items/prompt) were significantly related to mEMA compliance. For prompt frequency, the overall model was nonsignificant (*P*=.07), but the coefficient was significant (*P*<.001). Prompting 1 to 3 times per day was associated with higher compliance (87%; 95% CI 82.5-90.4) compared with studies with more than 3 prompts per day (76.9%) and 6 or more prompts per day (79.4%). The number of items per prompt was significant for both the overall model (*P*=.04) and the coefficient (*P*<.001). Factor analysis showed that prompts with more than 26 items had significantly lower compliance (63%; 95% CI 42.3-79.7) compared with prompts with ≤26 items (categories: ≤5; >5 to ≤9; >9.5 to ≤26; compliance range: 84%-78.6%).

For clinical data sets (n=27), no factors were significantly related to compliance. The number of items per prompt approached significance (*P*=.05). Compliance appeared to be lower in studies with 9.5-26 items per prompt (71.1%; 95% CI 62.5-78.6). Significant heterogeneity was reported for all significant findings (nonclinical and clinical), with I^2 values in excess of 90%, suggesting that although some variance can be explained by the significant factors, a large amount of variance remained unexplained. The burden score was not significantly related to compliance. The meta-analysis factor analysis compliance proportions are presented in Table 2.

Table 2. Meta-analysis results for clinical and nonclinical data sets.

Characteristics	Clinical data sets, n=27		Nonclinical data sets, n=41	
Protocol factors	n (%)	Pooled compliance (95% CI)	n (%)	Pooled compliance (95% CI)
Monitoring period, day				
<7	12 (44)	81.6 (74.1-87.3)	24 (58)	77.4 (71.3-85.5)
>7 to ≤14	4 (15)	84.4 (74.3-91.1)	9 (22)	82.1 (71.30-89.5)
>14	11 (41)	86.7 (81.2-91.0)	8 (19)	85.3 (80.5-89.1)
Device[a]				
Mobile	5 (19)	88.6 (71.5-96.1)	17 (41)	78.6 (71.9-84.0)
PDA	18 (66)	81.9 (77.4-85.8)	22 (54)	80.2 (74.2-84.9)
Other	4 (15)	88.8 (82.4-93.1)	2 (5)	92.2 (86.3-95.7)
Training				
Yes	23 (85)	84.4 (79.7-88.4)	36 (88)	80.4 (76.0-84.3)
No	0 (0)	N/A[b]	0 (0)	N/A
NR[c]	4 (15)	82.8 (78.4-86.4)	6 (15)	77.7 (73.1-82.0)
Incentives				
Yes	13 (48)	83.6 (77.7-88.3)	35 (85)	80.4 (79.0-84.3)
No	0 (0)	N/A	6 (15)	77.9 (73.1-82.0)
NR	18 (66)	85.7 (81.3-89.3)	0 (0)	N/A
Prompt frequency, per day				
1-3	8 (30)	85.3 (77.6-90.7)	8 (19)	87.0 (82.5-90.4)
4-5	12 (44)	81.5 (75.8-85.9)	16 (39)	76.9 (70.1-82.5)
≥6	6 (22)	86.3 (74.1-92.4)	17 (41)	79.4 (71.1-85.5)
UTD[d]	1 (4)	90.6 (N/A)	0 (0)	N/A
Burden score				
0-283.5	4 (15)	86.2 (76.9-92.4)	11 (27)	80.5 (75.7-84.6)
284-810	7 (26)	86.4 (75.4-93.0)	10 (24)	79.6 (73.7-84.7)
811-1806	3 (11)	88.8 (64.8-97.1)	13 (31)	82.8 (73.7-89.1)
≥1807	7 (26)	85.3 (80.5-89.0)	4 (10)	79.1 (51.5-93.1)
Items per prompt				
<5	8 (30)	87.2 (80.7-91.9)	10 (24)	82.8 (77.2-87.2)
5 to ≤9.5	7 (26)	88.4 (76.9-94.6)	8 (19)	78.6 (67.5-86.8)
9.5 to ≤26	2 (7)	71.1 (62.5-78.6)	16 (39)	84.0 (79.0-88.0)
>26	6 (22)	87.2 (82.9-90.7)	4 (10)	63.0 (42.3-79.7)
NR	5 (19)	72.7 (68.4-76.9)	3 (7)	70.3 (40.4-89.2)
Number of prompt types				
1	18 (66)	82.6 (78.1-86.5)	25 (61)	79.6 (75.0-83.5)
2	6 (22)	86.4 (71.3-94.2)	11 (27)	83.3 (71.7-90.9)
3	3 (11)	87.2 (85.5-88.8)	5 (12)	77.7 (65.7-86.5)

[a]Device type included with categories: Mobile phone (total n=22; smartphone: clinical n=1; nonclinical n=14; mobile: clinical n=4, nonclinical n=3); PDA (total n=45; clinical n=22, nonclinical n=23); Other (total n=6; electronic diary: clinical n=2, nonclinical n=1; iPod: clinical n=1, nonclinical n=1; watch device: clinical n=1).

[b]N/A: not applicable.

[c]NR: not reported.

[d]UTD: unable to be determined.

Discussion

Principal Findings

This systematic review of observational studies aimed to describe protocols and compliance with mEMA for self-reported health-related behaviors and psychological constructs in adults. Across 105 unique data sets, the key findings of this review were as follows: (1) a variety of health-related behaviors and psychological constructs were assessed, with affect being the most common mEMA target; (2) mEMA protocols varied widely across studies; (3) compliance was inconsistently reported across studies; (4) meta-analysis estimated an overall compliance rate of 81.9% (95% CI 79.1-84.4), with no significant difference between nonclinical and clinical data sets or estimates before or after data exclusions; (5) compliance was associated with prompts per day and items per prompt (nonclinical); and 6) no compelling relationship was identified between key features of mEMA protocols representing *burden* and mEMA compliance.

mEMA Use in Adults for Health-Related Behaviors and Psychological Constructs

The mEMA targets identified in this review reflect those reported in previous systematic reviews: affect/mood [7,12,14,15], cognitions [13], symptoms [15], eating or dietary behaviors [10,11], physical activity [10], and smoking or alcohol consumption [5,6]. Likewise, clinical populations identified in this review (psychiatric or mental health conditions, chronic pain and fibromyalgia, eating disorders, and substance use) were generally consistent with those reported previously [5,7,11,12,14-16]. However, there were chronic conditions unique to this review: oral or dental health, cancer, stroke and traumatic brain injury (for each n=3, 9/41, 22%), HIV, and upper abdominal surgery (for each n=1, 2/41, 5%). The small number of studies identified for these clinical groups may suggest that the potential for mEMA has not yet been realized in these populations.

Reporting of mEMA Protocols and Compliance

Most studies included in this review provided information around the EMA protocol used (device, monitoring duration, frequency and type of prompts, provision of training, and use of incentives). Consistent with previous systematic reviews of both youth and adults, there was considerable heterogeneity across studies for EMA protocols (Multimedia Appendix 3). Heterogeneity may be expected given the various potential applications of this survey approach. The mEMA protocol required to obtain sufficient or appropriate self-reported data on daily habitual behaviors in the general population is not likely to be the same as that for obtaining self-reported data on psychological responses to events or stimuli in clinical contexts. For example, the average EMA monitoring duration for studies of nonclinical adults in this review was 7 days (range: 1-49 days) compared with 12 days (range: 1-182 days) for clinical populations and 30 days (range: 3-730 days) in a review of EMA in substance users [5]. Likewise, prompt type, frequency, and complexity are expected to differ depending on the EMA target and population. Reviews of studies of EMA for diet and physical activity (common behaviors) report a daily average prompt frequency of 20 [10] compared with less than 4 prompts per day in substance use [5]. For these reasons, in systematic reviews of EMA use—including this one—reporting of summary metrics (mean, SD, median, range) for protocol components could be interpreted as a reflection of diversity in EMA application rather than a lack of protocol standardization.

The same rationale cannot be applied to the inconsistencies identified in reporting of EMA protocol compliance. Compliance is problematic to determine for event-based prompts (eg, those completed with smoking or consumption of alcohol). Compliance for time-based notifications, especially when the EMA is conducted using mobile devices, is relatively simple (number of prompts answered out of the total number of prompts delivered). However, participants may respond to a notification but may not complete all survey items or may not respond in a timely manner, affecting the momentary aspect of the EMA. In both of these cases, the act of responding might appropriately contribute to compliance rates, but the data are unlikely to be valid. These concepts were evident in the earliest recommendations for reporting compliance in EMA studies [17], which predate the sampling frame of this systematic review (2006-2016 inclusive). Considering that 71 studies were excluded from this review because of the absence of reporting mEMA compliance, less than half of the studies included in this review complied with recommendations put forward by Stone and Shiffman [17], such as reporting the proportion of delivered prompts answered (43/105, 41%) or defining a criterion for valid EMA data (37/105, 35%). Similarly, less than half of the data sets included in this review reported an average number of prompts answered per person (44/105, 42%), as recommended by more recently published guidelines for reporting EMA [8,10].

With the growth of systematic review methodologies (meta-synthesis, meta-regression, etc), one aspect of reporting for EMA warrants further consideration. EMA allows collection of self-report data across multiple survey items reflecting a range of behavioral, psychological, and contextual factors. It is not uncommon for data collected in the original, primary study to be reported in several publications. The foci of these *offspring* publications may include the total original sample of participants recruited (eg, unpublished data for specific mEMA items or other variables) or explore a subset of the original study participants (eg, patterns associated with participant characteristics). Although this is a reasonable and defensible use of the original study's resources, identification of duplicate or overlapping data in studies can be problematic. Where ambiguity exists, contacting the study authors is one way to clarify which publication should be considered the primary report (and which report overlapping or duplicate data). However, this option becomes less practical as time and people move on. The alternative is for authors to include an explicit statement concerning the existence of publications that include overlapping or duplicate data. There were a number of exemplars of this aspect of reporting in studies included [67,68,96] and excluded from this review [115-118].

Associations Between Key Components of mEMA Protocols and Compliance: Meta-analysis

In our meta-analysis (68 data sets), which replicates and was guided by the authors of 2 previous meta-analyses on this topic [5,9], the overall compliance rate was 81.9% (95% CI 79.1-84.4). This was slightly higher than that reported by Wen et al [9] (78.26%; 95% CI 75.49-80.78) and Jones et al [5] (75.06%; 95% CI 72.37-77.65). Although concerns have been expressed about the relationship between EMA burden and compliance, it remains unclear whether, or which, EMA protocol factors affect participant compliance. In our meta-analysis, for nonclinical data sets, prompt frequency per day and the number of items per prompt were significantly related to compliance (noting that it is not unusual for coefficients derived within a model to be significant even when the overall model is not). However, the findings are likely affected by the number of data sets in some categories. For nonclinical data sets, frequencies of 1-3 prompts per day were associated with small but significantly higher mean cohort compliance. Higher compliance with lower number of prompts perhaps seems intuitive, yet the evidence is inconsistent. Wen et al [9] reported opposite patterns of significance when nonclinical and clinical population data were investigated, and Jones et al [5] and Ono et al [119] reported no relationship with prompt frequency and compliance among substance users and those affected by chronic pain, respectively.

The relationship between the number of items included within each prompt and compliance has not been explored in previous systematic reviews or meta-analyses of mEMA. In this review, the number of items respondents were required to complete in a standard prompt ranged from 1 to 73 (median 10), with a greater number of items more common in the mEMA of psychological constructs (Multimedia Appendix 3). Our analysis showed an intuitive relationship with compliance among nonclinical data (ie, ≥26 items per prompt had the lowest mean cohort compliance of 63%; 95% CI 42.3-79.7), but not with clinical data.

When aiming to identify protocol factors affecting compliance, the inconsistencies in reporting of EMA compliance and the likely publication bias (studies with lower compliance rates may not be submitted or accepted for publication) must also be considered [5]. These factors limit the inclusion of potentially eligible studies in meta-analyses (68/105, 64.8% data sets in this review; 36/42, 86% studies in a previous review [9]). In addition, studies included in meta-analyses privilege *best compliers* through exclusion of participants not meeting criteria for valid EMA data or compliance thresholds (determined a priori or posteriori). Jones et al [5] attempted to address this latter point by exploring protocol factors associated with participant data exclusions (monitoring duration and prompt frequency). Finally, aggregate level compliance may not be sensitive enough or provide sufficient resolution to identify factors associated with higher or lower compliance. While accepting these caveats, there are 2 ways to consider the results of the 3 meta-analyses undertaken by Wen et al [9], Jones et al [5], and this study:

1. There is insufficient resolution to identify associations—if they exist—at the aggregate data level.
2. Although confidence limits might be reduced by adding further studies, the meta-analyses are essentially correct, and the notion of protocol burden imposed on participants has little to no impact on compliance [4,5].

In studies using EMA, the issue of what constitutes an acceptable rate of compliance or missing data is debatable. Although several studies included in this review cite a criterion or commonly used threshold of 80%, we, similar to Jones et al [5], could not identify the derivation of this criterion. For authors currently planning, conducting, or writing papers or protocols on EMA to monitor health-related behaviors of psychological constructs, adequate recording and reporting of compliance data following recommendations by Liao et al [10] and Heron et al [8] should enable future meta-analyses to explore protocol factors affecting participant compliance rates.

This systematic review prospectively aimed to sample a decade of mEMA use (protocol registered in November 2016; sampling frame of 2006 to 2016) in observational studies including adults from clinical and nonclinical populations. As one of the first EMA reporting documents was published in 2002 [17], this sampling frame assumed that researchers planning or reporting studies including mEMA would be aware of these reporting recommendations. The time frame required for the uptake of EMA reporting recommendations is unknown, although estimates of the time required for uptake of translational research ranges between 2 and 17 years [120]. Our sampling frame and review, however, does not capture studies published from 2017 to date. It is possible that more recent publications differ from those included in our review (greater mobile phone use, better reporting of mEMA schedules, and compliance).

There are no universally accepted recommendations concerning the updating of systematic searches or incorporation of the newer studies into the review results. Systematic reviews—depending on the specific question and volume of studies eligible for inclusion—are time- and labor-intensive. For larger reviews, it is not uncommon for these to take >2 years [121], with updates of Cochrane Collaboration systematic reviews taking up to 3.3 years [122]. The current Cochrane Collaboration policy infers that the decision to update a systematic review should consider the importance of the review question and the volume of new information (studies) [122]. Early in the review process (postsearch completion), 2 papers were identified, published in 2016 [10] and 2017 [8], providing updated recommendations for EMA reporting. Although the volume of mEMA studies published from 2017 is substantial and growing, we opted not to undertake an updated search/meta-analysis to *quarantine* mEMA studies published before the availability of the more recent EMA reporting recommendations.

Strengths and Limitations

This review was strengthened by the broad eligibility criteria used, including studies across nonclinical and clinical contexts in adults. The meta-analysis method was replicated from previous studies [5,9], enabling direct comparison of findings. To the best of the authors' knowledge, this review is the first to propose and explore *burden* as a compound effect of the

various EMA factors (monitoring duration, prompt frequency and prompt type, item per prompt) on participant compliance. We have proposed this novel metric as a starting point for conversations, critique, and further development. In its current form, the burden metric does not include all factors likely to contribute to burden (unfamiliarity with technology, adjunctive use of wearable technologies such as accelerometers), the proposed weighting is rudimentary, and the accuracy of study design features was not confirmed by the study authors.

Limitations of this review include a search strategy focused on the use of mEMA and excluding interventions delivered using EMA (EMI). Consequently, the findings of this review should not be extrapolated or assumed to be similar in studies using EMI. Most studies included in this review provided a clear statement of the primary outcome of interest within each observational study, and we are confident that our categorization of primary mEMA targets is defensible. However, when observational studies did not clearly identify or infer a primary outcome of interest and given mEMA survey items can include multiple items for both self-reported behavioral and psychological constructs, for a small number of studies, misclassification may exist with respect to categorization of mEMA targets as primary or secondary. In the absence of explicit statements by the authors on the number of items within each standard notification, we adopted a conservative approach by estimating the minimum compulsory number of items based on either the information provided by authors within publications or reviewing the instruments reported by authors for inclusion within surveys. The impact of including only studies published in English is unknown.

Conclusions

This review suggests that there is substantial interest in the use of mEMA in adults to collect self-reported health-related behavior and psychological construct data in nonclinical and clinical contexts. Across mEMA studies, there was considerable heterogeneity in protocol design, which may reflect a concerted effort by researchers to tailor mEMA protocols for the intended target and/or population to facilitate compliance. However, the number of studies reporting participant compliance with EMA is concerning. As a result of no or underreporting of compliance, pooled compliance rates may be skewed in favor of overall higher EMA compliance rates. This may dampen associations between compliance rates and EMA protocol factors or burden, making it difficult to ascertain which, if any, protocol factors (such as prompt frequency and number of items within prompts, as identified in this analysis) improve compliance and data collection.

Acknowledgments

The authors sincerely thank Dr Cheng Wen and Dr Andrew Jones for their guidance regarding the meta-analysis process. This study was supported by the University of South Australian High Achiever Vacation scholarship scheme (authors AG and JI). This study was not sponsored.

Authors' Contributions

All authors contributed to this systematic review through the initiation and development of the original protocol (MW, HL, and FF), search and screening (AG and JI), data extraction (AG, JI, MW, HL, FF, and KF), synthesis and meta-analysis (KF, HL, FF, and MW), manuscript development, and final review (MW, HL, KF, FF, AG, and JI).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Ecological momentary assessment (EMA) population and compliance characteristics for studies included within review.
[DOCX File , 81 KB - jmir_v23i3e17023_app1.docx]

Multimedia Appendix 2
Summary of mobile ecological momentary assessment (mEMA) targets and participant characteristics in nonclinical and clinical mEMA studies.
[DOCX File , 67 KB - jmir_v23i3e17023_app2.docx]

Multimedia Appendix 3
Summary of mobile ecological momentary assessment protocols.
[DOCX File , 63 KB - jmir_v23i3e17023_app3.docx]

References

1. Burke LE, Shiffman S, Music E, Styn MA, Kriska A, Smailagic A, et al. Ecological momentary assessment in behavioral research: addressing technological and human participant challenges. J Med Internet Res 2017 Mar 15;19(3):e77 [FREE Full text] [doi: 10.2196/jmir.7138] [Medline: 28298264]

2. Dunton GF. Ecological momentary assessment in physical activity research. Exerc Sport Sci Rev 2017 Dec;45(1):48-54 [FREE Full text] [doi: 10.1249/JES.0000000000000092] [Medline: 27741022]
3. Rintala A, Wampers M, Myin-Germeys I, Viechtbauer W. Response compliance and predictors thereof in studies using the experience sampling method. Psychol Assess 2019 Feb;31(2):226-235. [doi: 10.1037/pas0000662] [Medline: 30394762]
4. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
5. Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IH, Wen CK, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019 Apr;114(4):609-619. [doi: 10.1111/add.14503] [Medline: 30461120]
6. Freedman MJ, Lester KM, McNamara C, Milby JB, Schumacher JE. Cell phones for ecological momentary assessment with cocaine-addicted homeless patients in treatment. J Subst Abuse Treat 2006 Mar;30(2):105-111. [doi: 10.1016/j.jsat.2005.10.005] [Medline: 16490673]
7. Baltasar-Tello I, Miguélez-Fernández C, Peñuelas-Calvo I, Carballo JJ. Ecological momentary assessment and mood disorders in children and adolescents: a systematic review. Curr Psychiatry Rep 2018 Aug 01;20(8):66. [doi: 10.1007/s11920-018-0913-z] [Medline: 30069650]
8. Heron KE, Everhart RS, McHale SM, Smyth JM. Using mobile-technology-based ecological momentary assessment (EMA) methods with youth: a systematic review and recommendations. J Pediatr Psychol 2017 Nov 01;42(10):1087-1107. [doi: 10.1093/jpepsy/jsx078] [Medline: 28475765]
9. Wen CKF, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017 Apr 26;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
10. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE checklist for reporting EMA studies (CREMAS). J Med Internet Res 2016 Jun 21;18(6):e151 [FREE Full text] [doi: 10.2196/jmir.4954] [Medline: 27328833]
11. Schembre SM, Liao Y, O'Connor SG, Hingle MD, Shen S, Hamoy KG, et al. Mobile ecological momentary diet assessment methods for behavioral research: systematic review. JMIR Mhealth Uhealth 2018 Nov 20;6(11):e11170 [FREE Full text] [doi: 10.2196/11170] [Medline: 30459148]
12. Miguelez-Fernandez C, de Leon SJ, Baltasar-Tello I, Peñuelas-Calvo I, Barrigon ML, Capdevila AS, et al. Evaluating attention-deficit/hyperactivity disorder using ecological momentary assessment: a systematic review. Atten Defic Hyperact Disord 2018 Dec;10(4):247-265. [doi: 10.1007/s12402-018-0261-1] [Medline: 30132248]
13. Moore RC, Swendsen J, Depp CA. Applications for self-administered mobile cognitive assessments in clinical research: a systematic review. Int J Methods Psychiatr Res 2017 Dec;26(4):e1562 [FREE Full text] [doi: 10.1002/mpr.1562] [Medline: 28370881]
14. Bos FM, Schoevers RA, aan het Rot M. Experience sampling and ecological momentary assessment studies in psychopharmacology: a systematic review. Eur Neuropsychopharmacol 2015 Nov;25(11):1853-1864. [doi: 10.1016/j.euroneuro.2015.08.008] [Medline: 26336868]
15. May M, Junghaenel DU, Ono M, Stone AA, Schneider S. Ecological momentary assessment methodology in chronic pain research: a systematic review. J Pain 2018 Jul;19(7):699-716. [doi: 10.1016/j.jpain.2018.01.006] [Medline: 29371113]
16. Bell IH, Lim MH, Rossell SL, Thomas N. Ecological momentary assessment and intervention in the treatment of psychotic disorders: a systematic review. Psychiatr Serv 2017 Nov 01;68(11):1172-1181. [doi: 10.1176/appi.ps.201600523] [Medline: 28669284]
17. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [Medline: 12173681]
18. Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. An evidence-based practice guideline for the peer review of electronic search strategies. J Clin Epidemiol 2009 Sep;62(9):944-952. [doi: 10.1016/j.jclinepi.2008.10.012] [Medline: 19230612]
19. Solhan MB, Trull TJ, Jahng S, Wood PK. Clinical assessment of affective instability: comparing EMA indices, questionnaire reports, and retrospective recall. Psychol Assess 2009 Sep;21(3):425-436 [FREE Full text] [doi: 10.1037/a0016869] [Medline: 19719353]
20. Sorbi MJ, Peters ML, Kruise DA, Maas CJ, Kerssens JJ, Verhaak PF, et al. Electronic momentary assessment in chronic pain I: psychological pain responses as predictors of pain intensity. Clin J Pain 2006 Jan;22(1):55-66. [doi: 10.1097/01.ajp.0000148624.46756.fa] [Medline: 16340594]
21. Johnson EI, Grondin O, Barrault M, Faytout M, Helbig S, Husky M, et al. Computerized ambulatory monitoring in psychiatry: a multi-site collaborative study of acceptability, compliance, and reactivity. Int J Methods Psychiatr Res 2009;18(1):48-57. [doi: 10.1002/mpr.276] [Medline: 19195050]
22. Kimhy D, Vakhrusheva J, Khan S, Chang RW, Hansen MC, Ballon JS, et al. Emotional granularity and social functioning in individuals with schizophrenia: an experience sampling study. J Psychiatr Res 2014 Jun;53:141-148 [FREE Full text] [doi: 10.1016/j.jpsychires.2014.01.020] [Medline: 24561000]

23. Ritz T, Rosenfield D, Steptoe A. Physical activity, lung function, and shortness of breath in the daily life of individuals with asthma. Chest 2010 Oct;138(4):913-918. [doi: 10.1378/chest.08-3073] [Medline: 20472861]
24. MacKerron G, Mourato S. Happiness is greater in natural environments. Glob Environ Change 2013 Oct;23(5):992-1000. [doi: 10.1016/j.gloenvcha.2013.03.010]
25. Trampe D, Quoidbach J, Taquet M. Emotions in everyday life. PLoS One 2015;10(12):e0145450 [FREE Full text] [doi: 10.1371/journal.pone.0145450] [Medline: 26698124]
26. Jean FA, Swendsen JD, Sibon I, Fehér K, Husky M. Daily life behaviors and depression risk following stroke: a preliminary study using ecological momentary assessment. J Geriatr Psychiatry Neurol 2013 Sep;26(3):138-143. [doi: 10.1177/0891988713484193] [Medline: 23584854]
27. Granholm E, Ben-Zeev D, Fulford D, Swendsen J. Ecological momentary assessment of social functioning in schizophrenia: impact of performance appraisals and affect on social interactions. Schizophr Res 2013 Apr;145(1-3):120-124 [FREE Full text] [doi: 10.1016/j.schres.2013.01.005] [Medline: 23402693]
28. Wray TB, Kahler CW, Monti PM. Using Ecological Momentary Assessment (EMA) to study sex events among very high-risk men who have sex with men (MSM). AIDS Behav 2016 Jan 8:2231-2242. [doi: 10.1007/s10461-015-1272-y] [Medline: 26746212]
29. Zawadzki MJ, Smyth JM, Costigan HJ. Real-time associations between engaging in leisure and daily health and well-being. Ann Behav Med 2015 Aug;49(4):605-615. [doi: 10.1007/s12160-015-9694-3] [Medline: 25724635]
30. Armey MF, Crowther JH, Miller IW. Changes in ecological momentary assessment reported affect associated with episodes of nonsuicidal self-injury. Behav Ther 2011 Dec;42(4):579-588. [doi: 10.1016/j.beth.2011.01.002] [Medline: 22035987]
31. Cook PF, McElwain CJ, Bradley-Springer LA. Feasibility of a daily electronic survey to study prevention behavior with HIV-infected individuals. Res Nurs Health 2010 Jun;33(3):221-234. [doi: 10.1002/nur.20381] [Medline: 20499392]
32. Kaplan SE, Ohrbach R. Self-report of waking-state oral parafunctional behaviors in the natural environment. J Oral Facial Pain Headache 2016;30(2):107-119 [FREE Full text] [doi: 10.11607/ofph.1592] [Medline: 27128474]
33. Hacker ED, Ferrans CE. Ecological momentary assessment of fatigue in patients receiving intensive cancer therapy. J Pain Symptom Manage 2007 Mar;33(3):267-275. [doi: 10.1016/j.jpainsymman.2006.08.007] [Medline: 17349496]
34. Chandra S, Scharf D, Shiffman S. Within-day temporal patterns of smoking, withdrawal symptoms, and craving. Drug Alcohol Depend 2011 Sep 01;117(2-3):118-125 [FREE Full text] [doi: 10.1016/j.drugalcdep.2010.12.027] [Medline: 21324611]
35. Yoshiuchi K, Cook DB, Ohashi K, Kumano H, Kuboki T, Yamamoto Y, et al. A real-time assessment of the effect of exercise in chronic fatigue syndrome. Physiol Behav 2007 Dec 05;92(5):963-968 [FREE Full text] [doi: 10.1016/j.physbeh.2007.07.001] [Medline: 17655887]
36. Thomas JG, Doshi S, Crosby RD, Lowe MR. Ecological momentary assessment of obesogenic eating behavior: combining person-specific and environmental predictors. Obesity (Silver Spring) 2011 Aug;19(8):1574-1579 [FREE Full text] [doi: 10.1038/oby.2010.335] [Medline: 21273995]
37. Ottaviani C, Medea B, Lonigro A, Tarvainen M, Couyoumdjian A. Cognitive rigidity is mirrored by autonomic inflexibility in daily life perseverative cognition. Biol Psychol 2015 Apr;107:24-30. [doi: 10.1016/j.biopsycho.2015.02.011] [Medline: 25749107]
38. Kuntsche E, Labhart F. ICAT: development of an internet-based data collection method for ecological momentary assessment using personal cell phones. Eur J Psychol Assess 2013;29(2):140-148 [FREE Full text] [doi: 10.1027/1015-5759/a000137] [Medline: 24285917]
39. Ambwani S, Roche MJ, Minnick AM, Pincus AL. Negative affect, interpersonal perception, and binge eating behavior: an experience sampling study. Int J Eat Disord 2015 Sep;48(6):715-726. [doi: 10.1002/eat.22410] [Medline: 25946681]
40. Andersson C, Söderpalm Gordh AH, Berglund M. Use of real-time interactive voice response in a study of stress and alcohol consumption. Alcohol Clin Exp Res 2007 Nov;31(11):1908-1912. [doi: 10.1111/j.1530-0277.2007.00520.x] [Medline: 17949395]
41. Berg KC, Peterson CB, Crosby RD, Cao L, Crow SJ, Engel SG, et al. Relationship between daily affect and overeating-only, loss of control eating-only, and binge eating episodes in obese adults. Psychiatry Res 2014 Jan 30;215(1):185-191 [FREE Full text] [doi: 10.1016/j.psychres.2013.08.023] [Medline: 24200217]
42. Buckner JD, Crosby RD, Wonderlich SA, Schmidt NB. Social anxiety and cannabis use: an analysis from ecological momentary assessment. J Anxiety Disord 2012 Mar;26(2):297-304 [FREE Full text] [doi: 10.1016/j.janxdis.2011.12.006] [Medline: 22246109]
43. Burt SA, Donnellan MB. Evidence that the subtypes of antisocial behavior questionnaire (STAB) predicts momentary reports of acting-out behaviors. Pers Individ Dif 2010 Jun;48(8):917-920. [doi: 10.1016/j.paid.2010.02.021]
44. Businelle MS, Ma P, Kendzor DE, Reitzel LR, Chen M, Lam CY, et al. Predicting quit attempts among homeless smokers seeking cessation treatment: an ecological momentary assessment study. Nicotine Tob Res 2014 Oct;16(10):1371-1378 [FREE Full text] [doi: 10.1093/ntr/ntu088] [Medline: 24893602]
45. Clasen PC, Fisher AJ, Beevers CG. Mood-reactive self-esteem and depression vulnerability: person-specific symptom dynamics via smart phone assessment. PLoS One 2015;10(7):e0129774 [FREE Full text] [doi: 10.1371/journal.pone.0129774] [Medline: 26131724]

46. Courvoisier DS, Eid M, Lischetzke T, Schreiber WH. Psychometric properties of a computerized mobile phone method for assessing mood in daily life. Emotion 2010 Feb;10(1):115-124. [doi: 10.1037/a0017813] [Medline: 20141308]
47. Doherty ST, Lemieux CJ, Canally C. Tracking human activity and well-being in natural environments using wearable sensors and experience sampling. Soc Sci Med 2014 Apr;106:83-92. [doi: 10.1016/j.socscimed.2014.01.048] [Medline: 24549253]
48. Fitzsimmons-Craft EE, Bardone-Cone AM, Crosby RD, Engel SG, Wonderlich SA, Bulik CM. Mediators of the relationship between thin-ideal internalization and body dissatisfaction in the natural environment. Body Image 2016 Sep;18:113-122 [FREE Full text] [doi: 10.1016/j.bodyim.2016.06.006] [Medline: 27391791]
49. Fonareva I, Amen AM, Ellingson RM, Oken BS. Differences in stress-related ratings between research center and home environments in dementia caregivers using ecological momentary assessment. Int Psychogeriatr 2012 Jan;24(1):90-98 [FREE Full text] [doi: 10.1017/S1041610211001414] [Medline: 21777503]
50. Heron KE, Scott SB, Sliwinski MJ, Smyth JM. Eating behaviors and negative affect in college women's everyday lives. Int J Eat Disord 2014 Dec;47(8):853-859 [FREE Full text] [doi: 10.1002/eat.22292] [Medline: 24797029]
51. Heron KE, Smyth JM. Body image discrepancy and negative affect in women's everyday lives: an ecological momentary assessment evaluation of self-discrepancy theory. J Soc Clin Psychol 2013 Mar;32(3):276-295. [doi: 10.1521/jscp.2013.32.3.276]
52. Hofmann W, Wisneski DC, Brandt MJ, Skitka LJ. Morality in everyday life. Science 2014 Sep 12;345(6202):1340-1343 [FREE Full text] [doi: 10.1126/science.1251560] [Medline: 25214626]
53. Juth V, Dickerson SS, Zoccola PM, Lam S. Understanding the utility of emotional approach coping: evidence from a laboratory stressor and daily life. Anxiety Stress Coping 2015;28(1):50-70 [FREE Full text] [doi: 10.1080/10615806.2014.921912] [Medline: 24804564]
54. Kashdan TB, Collins RL. Social anxiety and the experience of positive emotion and anger in everyday life: an ecological momentary assessment approach. Anxiety Stress Coping 2010 May;23(3):259-272. [doi: 10.1080/10615800802641950] [Medline: 19326272]
55. Kirchner TR, Cantrell J, Anesetti-Rothermel A, Ganz O, Vallone DM, Abrams DB. Geospatial exposure to point-of-sale tobacco: real-time craving and smoking-cessation outcomes. Am J Prev Med 2013 Oct;45(4):379-385 [FREE Full text] [doi: 10.1016/j.amepre.2013.05.016] [Medline: 24050412]
56. Komulainen E, Meskanen K, Lipsanen J, Lahti JM, Jylhä P, Melartin T, et al. The effect of personality on daily life emotional processes. PLoS One 2014;9(10):e110907 [FREE Full text] [doi: 10.1371/journal.pone.0110907] [Medline: 25343494]
57. Lange S, Süß H. Measuring slips and lapses when they occur - ambulatory assessment in application to cognitive failures. Conscious Cogn 2014 Feb;24:1-11. [doi: 10.1016/j.concog.2013.12.008] [Medline: 24384496]
58. Liao Y, Intille SS, Dunton GF. Using ecological momentary assessment to understand where and with whom adults' physical and sedentary activity occur. Int J Behav Med 2015 Feb;22(1):51-61. [doi: 10.1007/s12529-014-9400-z] [Medline: 24639067]
59. Ramirez J, Miranda R. Alcohol craving in adolescents: bridging the laboratory and natural environment. Psychopharmacology (Berl) 2014 Apr;231(8):1841-1851 [FREE Full text] [doi: 10.1007/s00213-013-3372-6] [Medline: 24363093]
60. Riediger M, Wrzus C, Schmiedek F, Wagner GG, Lindenberger U. Is seeking bad mood cognitively demanding? Contra-hedonic orientation and working-memory capacity in everyday life. Emotion 2011 Jun;11(3):656-665. [doi: 10.1037/a0022756] [Medline: 21534659]
61. Robertson BM, Piasecki TM, Slutske WS, Wood PK, Sher KJ, Shiffman S, et al. Validity of the hangover symptoms scale: evidence from an electronic diary study. Alcohol Clin Exp Res 2012;36(1):171-177. [doi: 10.1111/j.1530-0277.2011.01592.x] [Medline: 21762183]
62. Rowan PJ, Cofta-Woerpel L, Mazas CA, Vidrine JI, Reitzel LR, Cinciripini PM, et al. Evaluating reactivity to ecological momentary assessment during smoking cessation. Exp Clin Psychopharmacol 2007 Aug;15(4):382-389. [doi: 10.1037/1064-1297.15.4.382] [Medline: 17696685]
63. Rutledge T, Stucky E, Dollarhide A, Shively M, Jain S, Wolfson T, et al. A real-time assessment of work stress in physicians and nurses. Health Psychol 2009 Mar;28(2):194-200. [doi: 10.1037/a0013145] [Medline: 19290711]
64. Schuster RM, Mermelstein RJ, Hedeker D. Ecological momentary assessment of working memory under conditions of simultaneous marijuana and tobacco use. Addiction 2016 Aug;111(8):1466-1476. [doi: 10.1111/add.13342] [Medline: 26857917]
65. Seto E, Hua J, Wu L, Shia V, Eom S, Wang M, et al. Models of individual dietary behavior based on smartphone data: the influence of routine, physical activity, emotion, and food environment. PLoS One 2016;11(4):e0153085 [FREE Full text] [doi: 10.1371/journal.pone.0153085] [Medline: 27049852]
66. Setodji CM, Martino SC, Scharf DM, Shadel WG. Quantifying the persistence of pro-smoking media effects on college students' smoking risk. J Adolesc Health 2014 Apr;54(4):474-480 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.09.011] [Medline: 24268361]
67. Shiffman S, Balabanis MH, Gwaltney CJ, Paty JA, Gnys M, Kassel JD, et al. Prediction of lapse from associations between smoking and situational antecedents assessed by ecological momentary assessment. Drug Alcohol Depend 2007 Dec 1;91(2-3):159-168 [FREE Full text] [doi: 10.1016/j.drugalcdep.2007.05.017] [Medline: 17628353]

68. Simons JS, Emery NN, Simons RM, Wills TA, Webb MK. Effects of alcohol, rumination, and gender on the time course of negative affect. Cogn Emot 2017 Nov;31(7):1405-1418 [FREE Full text] [doi: 10.1080/02699931.2016.1226162] [Medline: 27609298]
69. Spook JE, Paulussen T, Kok G, Van EP. Monitoring dietary intake and physical activity electronically: feasibility, usability, and ecological validity of a mobile-based Ecological Momentary Assessment tool. J Med Internet Res 2013;15(9):e214 [FREE Full text] [doi: 10.2196/jmir.2617] [Medline: 24067298]
70. Thielsch C, Andor T, Ehring T. Do metacognitions and intolerance of uncertainty predict worry in everyday life? An ecological momentary assessment study. Behav Ther 2015 Jul;46(4):532-543. [doi: 10.1016/j.beth.2015.05.001] [Medline: 26163716]
71. Tiplady B, Oshinowo B, Thomson J, Drummond GB. Alcohol and cognitive function: assessment in everyday life and laboratory settings using mobile phones. Alcohol Clin Exp Res 2009 Dec;33(12):2094-2102. [doi: 10.1111/j.1530-0277.2009.01049.x] [Medline: 19740132]
72. Waters AJ, Szeto EH, Wetter DW, Cinciripini PM, Robinson JD, Li Y. Cognition and craving during smoking cessation: an ecological momentary assessment study. Nicotine Tob Res 2014 May;16 Suppl 2:S111-S118 [FREE Full text] [doi: 10.1093/ntr/ntt108] [Medline: 23901053]
73. Witkiewitz K, Desai SA, Steckler G, Jackson KM, Bowen S, Leigh BC, et al. Concurrent drinking and smoking among college students: an event-level analysis. Psychol Addict Behav 2012 Sep;26(3):649-654 [FREE Full text] [doi: 10.1037/a0025363] [Medline: 21895348]
74. Zenk SN, Horoi I, McDonald A, Corte C, Riley B, Odoms-Young AM. Ecological momentary assessment of environmental and personal factors and snack food intake in African American women. Appetite 2014 Dec;83:333-341. [doi: 10.1016/j.appet.2014.09.008] [Medline: 25239402]
75. Aaron LA, Turner JA, Mancl LA, Sawchuk CN, Huggins KH, Truelove EL. Daily pain coping among patients with chronic temporomandibular disorder pain: an electronic diary study. J Orofac Pain 2006;20(2):125-137. [Medline: 16708830]
76. Dewey D, McDonald MK, Brown WJ, Boyd SJ, Bunnell BE, Schuldberg D. The impact of ecological momentary assessment on posttraumatic stress symptom trajectory. Psychiatry Res 2015 Dec 15;230(2):300-303. [doi: 10.1016/j.psychres.2015.09.009] [Medline: 26381184]
77. Dhingra LK, Homel P, Grossman B, Chen J, Scharaga E, Calamita S, et al. Ecological momentary assessment of smoking behavior in persistent pain patients. Clin J Pain 2014 Mar;30(3):205-213. [doi: 10.1097/AJP.0b013e31829821c7] [Medline: 23689351]
78. Epstein DH, Tyburski M, Craig IM, Phillips KA, Jobes ML, Vahabzadeh M, et al. Real-time tracking of neighborhood surroundings and mood in urban drug misusers: application of a new method to study behavior in its geographical context. Drug Alcohol Depend 2014 Jan 1;134:22-29 [FREE Full text] [doi: 10.1016/j.drugalcdep.2013.09.007] [Medline: 24332365]
79. Fitzsimmons-Craft EE, Accurso EC, Ciao AC, Crosby RD, Cao L, Pisetsky EM, et al. Restrictive eating in anorexia nervosa: examining maintenance and consequences in the natural environment. Int J Eat Disord 2015 Nov;48(7):923-931 [FREE Full text] [doi: 10.1002/eat.22439] [Medline: 26310991]
80. Hachizuka M, Yoshiuchi K, Yamamoto Y, Iwase S, Nakagawa K, Kawagoe K, et al. Development of a personal digital assistant (PDA) system to collect symptom information from home hospice patients. J Palliat Med 2010 Jun;13(6):647-651. [doi: 10.1089/jpm.2009.0350] [Medline: 20509795]
81. Juengst SB, Graham KM, Pulantara IW, McCue M, Whyte EM, Dicianno BE, et al. Pilot feasibility of an mHealth system for conducting ecological momentary assessment of mood-related symptoms following traumatic brain injury. Brain Inj 2015 Aug;29(11):1351-1361. [doi: 10.3109/02699052.2015.1045031] [Medline: 26287756]
82. Kulich K, Keininger DL, Tiplady B, Banerji D. Symptoms and impact of COPD assessed by an electronic diary in patients with moderate-to-severe COPD: psychometric results from the SHINE study. Int J Chron Obstruct Pulmon Dis 2015;10:79-94 [FREE Full text] [doi: 10.2147/COPD.S73092] [Medline: 25609942]
83. Kuroi R, Minakuchi H, Hara ES, Kawakami A, Maekawa K, Okada H, et al. A risk factor analysis of accumulated postoperative pain and swelling sensation after dental implant surgery using a cellular phone-based real-time assessment. J Prosthodont Res 2015 Jul;59(3):194-198. [doi: 10.1016/j.jpor.2015.05.003] [Medline: 26077378]
84. Lavender JM, De Young KP, Wonderlich SA, Crosby RD, Engel SG, Mitchell JE, et al. Daily patterns of anxiety in anorexia nervosa: associations with eating disorder behaviors in the natural environment. J Abnorm Psychol 2013 Aug;122(3):672-683 [FREE Full text] [doi: 10.1037/a0031823] [Medline: 23647124]
85. Merwin RM, Dmitrieva NO, Honeycutt LK, Moskovich AA, Lane JD, Zucker NL, et al. Momentary predictors of insulin restriction among adults with type 1 diabetes and eating disorder symptomatology. Diabetes Care 2015 Nov;38(11):2025-2032 [FREE Full text] [doi: 10.2337/dc15-0753] [Medline: 26384389]
86. Munsch S, Meyer AH, Milenkovic N, Schlup B, Margraf J, Wilhelm FH. Ecological momentary assessment to evaluate cognitive-behavioral treatment for binge eating disorder. Int J Eat Disord 2009 Nov;42(7):648-657. [doi: 10.1002/eat.20657] [Medline: 19197978]
87. Okifuji A, Bradshaw DH, Donaldson GW, Turk DC. Sequential analyses of daily symptoms in women with fibromyalgia syndrome. J Pain 2011 Jan;12(1):84-93 [FREE Full text] [doi: 10.1016/j.jpain.2010.05.003] [Medline: 20591745]

88. Sohl SJ, Friedberg F. Memory for fatigue in chronic fatigue syndrome: relationships to fatigue variability, catastrophizing, and negative affect. Behav Med 2008;34(1):29-38 [FREE Full text] [doi: 10.3200/BMED.34.1.29-38] [Medline: 18400687]
89. Thielsch C, Ehring T, Nestler S, Wolters J, Kopei I, Rist F, et al. Metacognitions, worry and sleep in everyday life: studying bidirectional pathways using Ecological Momentary Assessment in GAD patients. J Anxiety Disord 2015 Jun;33:53-61. [doi: 10.1016/j.janxdis.2015.04.007] [Medline: 26005837]
90. Weaver A, Young AM, Rowntree J, Townsend N, Pearson S, Smith J, et al. Application of mobile phone technology for managing chemotherapy-associated side-effects. Ann Oncol 2007 Nov;18(11):1887-1892 [FREE Full text] [doi: 10.1093/annonc/mdm354] [Medline: 17921245]
91. Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. J Med Internet Res 2016 Mar 29;18(3):e72 [FREE Full text] [doi: 10.2196/jmir.5505] [Medline: 27025287]
92. Dunbar MS, Scharf D, Kirchner T, Shiffman S. Do smokers crave cigarettes in some smoking situations more than others? Situational correlates of craving when smoking. Nicotine Tob Res 2010 Mar;12(3):226-234 [FREE Full text] [doi: 10.1093/ntr/ntp198] [Medline: 20133379]
93. Hofmann W, Adriaanse M, Vohs KD, Baumeister RF. Dieting and the self-control of eating in everyday environments: an experience sampling study. Br J Health Psychol 2014 Sep;19(3):523-539 [FREE Full text] [doi: 10.1111/bjhp.12053] [Medline: 23751109]
94. Hughes CD, Gunthert K, Wenze S, German R. The subscale specificity of the Affective Control Scale: Ecological validity and predictive validity of feared emotions. Motiv Emot 2015 Jun 5;39(6):984-992. [doi: 10.1007/s11031-015-9497-7]
95. Huguet A, McGrath PJ, Wheaton M, Mackinnon SP, Rozario S, Tougas ME, et al. Testing the feasibility and psychometric properties of a mobile diary (myWHI) in adolescents and young adults with headaches. JMIR Mhealth Uhealth 2015 May 08;3(2):e39 [FREE Full text] [doi: 10.2196/mhealth.3879] [Medline: 25956377]
96. Kanning M, Hansen S. Need satisfaction moderates the association between physical activity and affective states in adults aged 50+: an activity-triggered ambulatory assessment. Ann Behav Med 2017 Feb;51(1):18-29 [FREE Full text] [doi: 10.1007/s12160-016-9824-6] [Medline: 27539030]
97. Kashdan TB, Farmer AS. Differentiating emotions across contexts: comparing adults with and without social anxiety disorder using random, social interaction, and daily experience sampling. Emotion 2014 Jun;14(3):629-638 [FREE Full text] [doi: 10.1037/a0035796] [Medline: 24512246]
98. Kwapil TR, Barrantes-Vidal N, Armistead MS, Hope GA, Brown LH, Silvia PJ, et al. The expression of bipolar spectrum psychopathology in daily life. J Affect Disord 2011 Apr;130(1-2):166-170 [FREE Full text] [doi: 10.1016/j.jad.2010.10.025] [Medline: 21056476]
99. Schüz B, Bower J, Ferguson SG. Stimulus control and affect in dietary behaviours. An intensive longitudinal study. Appetite 2015 Apr;87:310-317. [doi: 10.1016/j.appet.2015.01.002] [Medline: 25579222]
100. Schüz N, Walters JAE, Frandsen M, Bower J, Ferguson SG. Compliance with an EMA monitoring protocol and its relationship with participant and smoking characteristics. Nicotine Tob Res 2014 May;16(Suppl 2):S88-S92. [doi: 10.1093/ntr/ntt142] [Medline: 24052500]
101. Schwerdtfeger A, Eberhardt R, Chmitorz A, Schaller E. Momentary affect predicts bodily movement in daily life: an ambulatory monitoring study. J Sport Exerc Psychol 2010 Oct;32(5):674-693. [Medline: 20980710]
102. Warthen MW, Tiffany ST. Evaluation of cue reactivity in the natural environment of smokers using ecological momentary assessment. Exp Clin Psychopharmacol 2009 Apr;17(2):70-77 [FREE Full text] [doi: 10.1037/a0015617] [Medline: 19331483]
103. Vasconcelos E Sa D, Wearden A, Hartley S, Emsley R, Barrowclough C. Expressed Emotion and behaviourally controlling interactions in the daily life of dyads experiencing psychosis. Psychiatry Res 2016 Nov 30;245:406-413. [doi: 10.1016/j.psychres.2016.08.060] [Medline: 27611070]
104. Ebner-Priemer UW, Kuo J, Schlotz W, Kleindienst N, Rosenthal MZ, Detterer L, et al. Distress and affective dysregulation in patients with borderline personality disorder: a psychophysiological ambulatory monitoring study. J Nerv Ment Dis 2008 Apr;196(4):314-320. [doi: 10.1097/NMD.0b013e31816a493f] [Medline: 18414126]
105. Green KT, Dennis PA, Neal LC, Hobkirk AL, Hicks TA, Watkins LL, et al. Exploring the relationship between posttraumatic stress disorder symptoms and momentary heart rate variability. J Psychosom Res 2016 Mar;82:31-34 [FREE Full text] [doi: 10.1016/j.jpsychores.2016.01.003] [Medline: 26944396]
106. Mazure CM, Weinberger AH, Pittman B, Sibon I, Swendsen J. Gender and stress in predicting depressive symptoms following stroke. Cerebrovasc Dis 2014;38(4):240-246 [FREE Full text] [doi: 10.1159/000365838] [Medline: 25401293]
107. Kothari DJ, Davis MC, Yeung EW, Tennen HA. Positive affect and pain: mediators of the within-day relation linking sleep quality to activity interference in fibromyalgia. Pain 2015 Mar;156(3):540-546 [FREE Full text] [doi: 10.1097/01.j.pain.0000460324.18138.0a] [Medline: 25679472]
108. Burns JW, Gerhart JI, Bruehl S, Peterson KM, Smith DA, Porter LS, et al. Anger arousal and behavioral anger regulation in everyday life among patients with chronic low back pain: Relationships to patient pain and function. Health Psychol 2015 May;34(5):547-555 [FREE Full text] [doi: 10.1037/hea0000091] [Medline: 25110843]

109. Smyth JM, Wonderlich SA, Heron KE, Sliwinski MJ, Crosby RD, Mitchell JE, et al. Daily and momentary mood and stress are associated with binge eating and vomiting in bulimia nervosa patients in the natural environment. J Consult Clin Psychol 2007 Aug;75(4):629-638. [doi: 10.1037/0022-006X.75.4.629] [Medline: 17663616]
110. Ainsworth J, Palmier-Claus JE, Machin M, Barrowclough C, Dunn G, Rogers A, et al. A comparison of two delivery modalities of a mobile phone-based assessment for serious mental illness: native smartphone application vs text-messaging only implementations. J Med Internet Res 2013;15(4):e60 [FREE Full text] [doi: 10.2196/jmir.2328] [Medline: 23563184]
111. Tomiyama AJ, Mann T, Comer L. Triggers of eating in everyday life. Appetite 2009 Feb;52(1):72-82 [FREE Full text] [doi: 10.1016/j.appet.2008.08.002] [Medline: 18773931]
112. Luczak SE, Rosen IG, Wall TL. Development of a real-time repeated-measures assessment protocol to capture change over the course of a drinking episode. Alcohol Alcohol 2015 Mar;50(2):180-187 [FREE Full text] [doi: 10.1093/alcalc/agu100] [Medline: 25568142]
113. Monk RL, Heim D, Qureshi A, Price A. 'I have no clue what I drunk last night' using Smartphone technology to compare in-vivo and retrospective self-reports of alcohol consumption. PLoS One 2015;10(5):e0126209 [FREE Full text] [doi: 10.1371/journal.pone.0126209] [Medline: 25992573]
114. Santangelo P, Reinhard I, Mussgay L, Steil R, Sawitzki G, Klein C, et al. Specificity of affective instability in patients with borderline personality disorder compared to posttraumatic stress disorder, bulimia nervosa, and healthy controls. J Abnorm Psychol 2014 Feb;123(1):258-272 [FREE Full text] [doi: 10.1037/a0035619] [Medline: 24661176]
115. Piasecki TM, Cooper ML, Wood PK, Sher KJ, Shiffman S, Heath AC. Dispositional drinking motives: associations with appraised alcohol effects and alcohol consumption in an ecological momentary assessment investigation. Psychol Assess 2014 Jun;26(2):363-369 [FREE Full text] [doi: 10.1037/a0035153] [Medline: 24274049]
116. Goldschmidt AB, Wonderlich SA, Crosby RD, Engel SG, Lavender JM, Peterson CB, et al. Ecological momentary assessment of stressful events and negative affect in bulimia nervosa. J Consult Clin Psychol 2014 Feb;82(1):30-39 [FREE Full text] [doi: 10.1037/a0034974] [Medline: 24219182]
117. Trela CJ, Piasecki TM, Bartholow BD, Heath AC, Sher KJ. The natural expression of individual differences in self-reported level of response to alcohol during ecologically assessed drinking episodes. Psychopharmacology (Berl) 2016 Jun;233(11):2185-2195 [FREE Full text] [doi: 10.1007/s00213-016-4270-5] [Medline: 27037938]
118. Tomko RL, Solhan MB, Carpenter RW, Brown WC, Jahng S, Wood PK, et al. Measuring impulsivity in daily life: the momentary impulsivity scale. Psychol Assess 2014 Jun;26(2):339-349 [FREE Full text] [doi: 10.1037/a0035083] [Medline: 24274047]
119. Ono M, Schneider S, Junghaenel DU, Stone AA. What affects the completion of ecological momentary assessments in chronic pain research? An individual patient data meta-analysis. J Med Internet Res 2019 Feb 05;21(2):e11398 [FREE Full text] [doi: 10.2196/11398] [Medline: 30720437]
120. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. J R Soc Med 2011 Dec;104(12):510-520 [FREE Full text] [doi: 10.1258/jrsm.2011.110180] [Medline: 22179294]
121. Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. Syst Rev 2012;1:10 [FREE Full text] [doi: 10.1186/2046-4053-1-10] [Medline: 22587960]
122. Elkins MR. Updating systematic reviews. J Physiother 2018 Jan;64(1):1-3 [FREE Full text] [doi: 10.1016/j.jphys.2017.11.009] [Medline: 29289593]

Abbreviations

EMA: ecological momentary assessment
ES: effect size
ESS: effective sample size
mEMA: mobile ecological momentary assessment
mEMI: mobile ecological momentary intervention

Edited by G Eysenbach; submitted 12.11.19; peer-reviewed by M May, G Dunton, K Heron; comments to author 23.12.19; revised version received 01.03.20; accepted 31.10.20; published 03.03.21.

Please cite as:
Williams MT, Lewthwaite H, Fraysse F, Gajewska A, Ignatavicius J, Ferrar K
Compliance With Mobile Ecological Momentary Assessment of Self-Reported Health-Related Behaviors and Psychological Constructs in Adults: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(3):e17023
URL: https://www.jmir.org/2021/3/e17023
doi:10.2196/17023
PMID:33656451

Review

Acceptance and Use of Home-Based Electronic Symptom Self-Reporting Systems in Patients With Cancer: Systematic Review

Youmin Cho[1*], MSN, AGPCNP-BC; Huiting Zhang[1], DNP, MSN, AGPCNP-BC; Marcelline Ruth Harris[1], PhD; Yang Gong[2], MD, PhD; Ellen Lavoie Smith[1], PhD, MSN, AOCN; Yun Jiang[1*], MSc, PhD

[1]University of Michigan School of Nursing, Ann Arbor, MI, United States

[2]University of Texas Health Science Center at Houston School of Biomedical Informatics, Houston, TX, United States

[*]these authors contributed equally

Corresponding Author:
Yun Jiang, MSc, PhD
University of Michigan School of Nursing
400 North Ingalls
Ann Arbor, MI, 48109
United States
Phone: 1 734 763 3705
Email: jiangyu@umich.edu

Abstract

Background: Electronic symptom self-reporting systems (e-SRS) have been shown to improve symptoms and survival in patients with cancer. However, patient engagement in using e-SRS for voluntary symptom self-reporting is less optimal. Multiple factors can potentially affect patients' acceptance and engagement in using home-based e-SRS. However, such factors have not been fully explored in cancer populations.

Objective: The aim of this study is to understand the acceptance and use of home-based e-SRS by patients with cancer and identify associated facilitators and barriers.

Methods: PubMed, CINAHL, Scopus, and PsycINFO (January 2010 to March 2020) were searched using a combination of Medical Subject Headings (MeSH) terms and keywords such as symptom self-reporting, electronic/technology, cancer, and their synonyms. Included studies focused on the use of home-based e-SRS by patients with cancer and their families. Studies on patients' use of e-SRS in clinical settings only were excluded. Of the 3740 papers retrieved, 33 were included in the final review. Factors associated with patient acceptance and use of e-SRS were extracted and synthesized.

Results: Most e-SRS were web based (22/33, 66%) or mobile app based (9/33, 27%). The e-SRS initial acceptance, represented by patient enrollment rates, ranged from 40% (22/55) to 100% (100/100). High e-SRS acceptance was rated by 69% (59/85) to 77.6% (337/434) of the patients after they used the system. The e-SRS use, measured by patients' response rates to questionnaires (ranging from 1596/3521, 45.33% to 92%) or system log-on rates (ranging from 4/12, 33% to 99/100, 99%), declined over time in general patterns. Few studies (n=7) reported e-SRS use beyond 6 months, with the response rates ranging from 62% (40/64) to 85.1% (541/636) and the log-on rates ranging from 63.6% (103/162) to 77% (49/64). The availability of compatible devices and technical support, interactive system features, information accessibility, privacy, questionnaire quality, patient physical/psychosocial status, and age were associated with patient acceptance and use of home-based e-SRS.

Conclusions: Acceptance and use of home-based e-SRS by patients with cancer varied significantly across studies, as assessed by a variety of approaches. The lack of access to technology has remained a barrier to e-SRS adoption. Interactive system features and personalized questionnaires may increase patient engagement. More studies are needed to further understand patients' long-term use of home-based e-SRS behavior patterns to develop personalized interventions to support symptom self-management and self-reporting of patients with cancer for optimal health outcomes.

(J Med Internet Res 2021;23(3):e24638) doi:10.2196/24638

KEYWORDS

symptom; self report; telemedicine; technology; internet; mobile phone; patient preference; cancer; patient-reported outcomes

Introduction

Background

Patient-reported symptoms, as patient-reported outcomes, are directly reported by patients without any editing or interpretation by clinicians [1,2]. The importance of collecting patient-reported symptoms has been increasingly recognized in cancer care because patients with cancer often experience unpredictable subjective symptoms, such as severe nausea, fatigue, or pain, which can lead to unwarranted emergency room visits or hospital admissions [2-4]. Multiple studies have shown that clinicians are less reliable in identifying subjective symptoms than patients; clinicians are more likely to underestimate the severity of symptoms and sometimes overlook the patient's self-report [5]. Thus, collecting symptom information directly from patients with cancer is an important component of effective symptom management and improved quality of cancer care.

There is growing evidence for the use of electronic technology systems to collect patient-reported symptoms [5]. Electronic symptom self-reporting systems (e-SRS) have a variety of advantages compared with paper-and-pencil–based reporting formats, including fewer errors in data entry, less missing data, less burden in data management, faster access to data, increased potential for adopting alerts and notifications, and improved real-time patient-provider communications [6-8]. For example, one study found that persons using paper diaries for tracking pain reported a high level of fake compliance (90% of patients reported the use of paper diaries for pain tracking, but only 32% actually used), whereas the electronic diaries group demonstrated 99% validated compliance [9].

Informed by the chronic care model, patients with cancer and their families are expected to be in partnership with clinicians for joint management of the disease and related consequences to improve the quality of cancer care [10]. Remote symptom reporting using electronic technology systems outside cancer clinic settings, that is, using telehealth, play an increasingly significant role in this partnership [11]. Using home-based e-SRS, patients can report their signs and symptoms earlier than waiting for their next clinical visits, facilitating more efficient and effective symptom management [5,12]. Home-based e-SRS can be cost-effective because of the low cost of data collection using electronic surveys and timely identification and management of early symptoms before becoming severe [11,13]. Although many studies have collected patient-reported symptoms in clinical settings [14-16], research has found that patients with cancer usually report fewer and/or less severe symptoms during clinical visits than when self-reported in real time from home [17]. In addition, clinic-based reporting systems may not be optimal for patients receiving oral anticancer therapies, who often have less frequent clinical follow-up visits.

e-SRS has been shown to improve symptoms and survival in patients with cancer [12,18,19]. However, to achieve these benefits, patients' acceptance and voluntary use of home-based e-SRS are essential for establishing long-term benefits of symptom self-reporting [8,20]. A literature review of 33 e-SRS used in cancer care highlights that 70% of reporting systems were provided with in-clinic access [18]. To date, there has been no synthesized evaluation of what is known about voluntary use of home-based e-SRS by patients with cancer. Multiple personal and technical factors can potentially affect patients' acceptance and use of home-based e-SRS; however, such factors have not been fully explored among cancer populations.

Objectives

This study aims to explore acceptance and use of home-based e-SRS by patients with cancer and facilitators/barriers associated with acceptance and use of home-based e-SRS by patients with cancer.

Methods

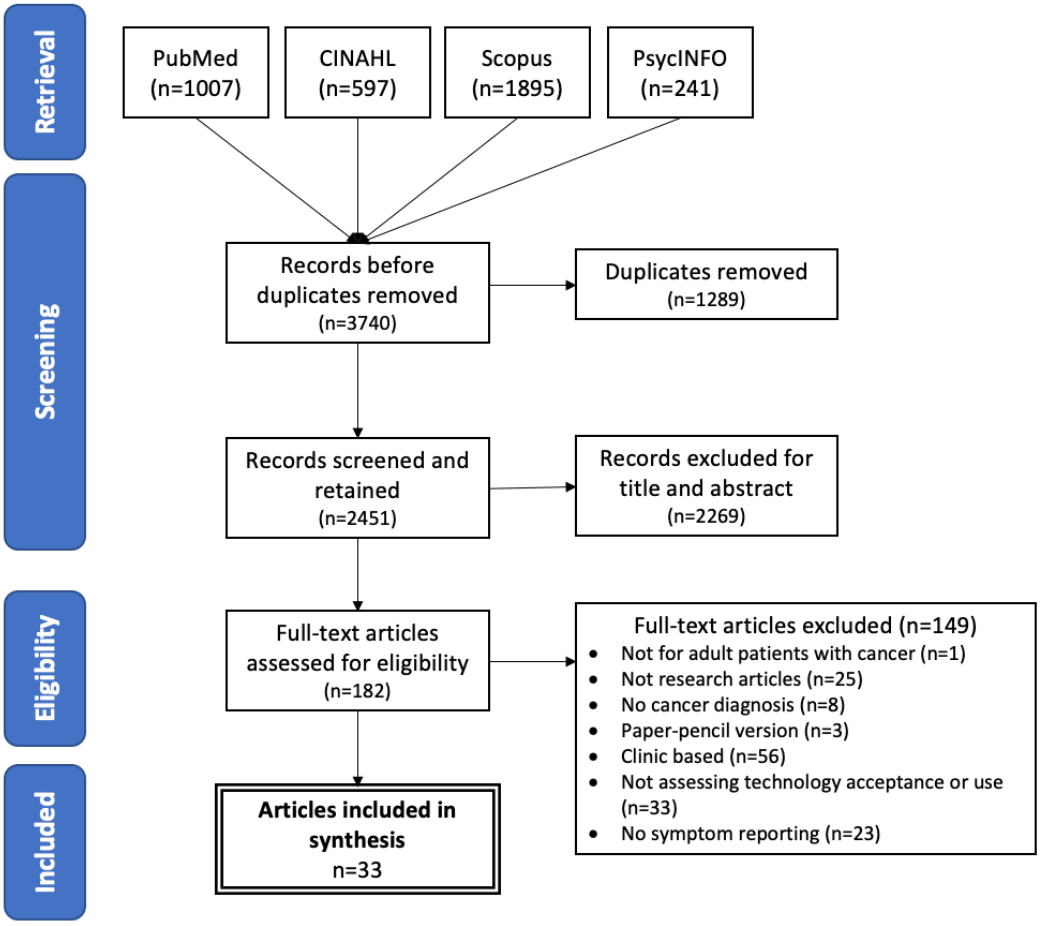
Search Strategy

Databases, including PubMed, CINAHL, Scopus, and PsycINFO, were searched for papers published between January 2010 and March 2020. A total of 3 groups of search terms—symptom self-reporting, electronic/technology, and cancer/oncology—were used in combination with their Medical Subject Headings (MeSH) terms, keywords, and synonyms. Synonyms were generated based on preliminary searches and some entry terms of MeSH terms (search strategies are included in Multimedia Appendix 1). We included papers that (1) included patients diagnosed with cancer who were aged ≥18 years, (2) reported patients or family members' use of an electronic version of symptom self-reporting systems/tools for symptom self-reporting outside of clinic or hospital settings, and (3) were original peer-reviewed research papers that were written in English. Studies published before 2010 were excluded because smartphones and tablets were not widely used until 2010. We excluded papers that reported the use of paper-based symptom self-reporting tools or clinic-based e-SRS only. Other excluded papers were those that did not provide measures or results specifically about patients' acceptance or use of home-based e-SRS or did not focus on symptom reporting.

Selection of Papers

A total of 3740 papers were retrieved from database searches. After removing duplicates and reviewing titles and abstracts for relevance, 182 papers remained for the full-text screening. Among them, 149 papers did not meet the inclusion criteria and were excluded, including not for adults (n=1), no cancer diagnosis (n=8), not research papers (n=25), used paper-pencil version of symptom reporting (n=3), not assessing acceptance and use (n=33), not symptom reporting (n=23), and clinic-based systems (n=56). A total of 33 papers were included in the final review. Figure 1 shows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart describing the overall search and selection process.

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



Data Extraction and Analysis

Study characteristics and information regarding e-SRS were extracted from each reviewed paper. Information regarding e-SRS acceptance and use was extracted from user surveys and postintervention interviews. Technology *acceptance* and *use* were defined based on the widely adopted Unified Theory of Acceptance and Use of Technology (UTAUT) [21]. Specifically, in this study, e-SRS acceptance was defined as patients' intention to use home-based e-SRS. Among the studies that did not directly assess patients' intention to use e-SRS before their actual use of the system, e-SRS initial acceptance was operationalized as patients' willingness to participate in the study using home-based e-SRS (eg, participant enrollment rate: the rate of enrollees out of all approached eligible patients) [22]. The e-SRS use was defined as the actual use rate of e-SRS or the description of patients' e-SRS use behavior. The e-SRS use rates were extracted and summarized based on the calculations reported in the studies, categorized as long-term (≥6 months) and short-term (<6 months) use. Potential facilitators/barriers to e-SRS acceptance and use were extracted and synthesized based on reported reasons for nonparticipation, users' feedback surveys, and postintervention interviews.

Critical Appraisal for Quality of Studies

The quality of studies was assessed using the Methodological Index for Nonrandomized Studies (MINORS), which includes 8 items for assessing noncomparative nonrandomized studies and 4 optional items for comparative studies (global Cronbach α=.73) [23]. As this study focused on users' acceptance and use of e-SRS, if a comparative study was reported, only the information from the technology user groups were analyzed. Therefore, this study only adopted the first 8 criteria of MINORS, including whether the study has a clear aim, clear inclusion and exclusion criteria, prospective data collection, appropriate endpoints, unbiased assessment, adequate study period, reasonable proportion of follow-up loss, and prospective sample size calculation [23]. Furthermore, the Critical Appraisal Skills Programme (CASP) Qualitative Research Checklist was used to evaluate the quality of a qualitative study or the qualitative design of a mixed methods study [24]. This 10-item checklist assesses the appropriateness of qualitative methodology, design, data collection, and analysis process, and the value of findings [24]. Items from both the MINORS and CASP checklist were graded on a scale of 0 (not reported), 1 (reported but inadequate), and 2 (reported and adequate). Studies with a MINORS score of 11 (out of a total score of 16) or less, or a CASP checklist score less than 15 (out of 20) were classified as low-quality studies [25,26]. Overall, both MINORS and CASP scores indicated adequate quality of the reviewed studies. The mean MINORS score was 13.6 (SD 1.4), with a range of 10-16 out of 16. Only 2 studies had a low-quality score below 11, mainly due to small sample sizes (n=5-21) and inadequate description of the study endpoints [27,28]. The mean

CASP checklist score was 17.7 (SD 1.5) out of 20 (range, 15-20). No study had a low-quality score below 15 (Multimedia Appendix 2).

Results

Summary of Study Characteristics

Among the reviewed papers, most studies (1) were conducted in the United States (15/33, 45%) or Europe (16/33, 48%), (2) recruited participants from tertiary cancer centers (25/33, 76%), (3) had sample sizes ranging from 5 to 3521, and (4) reported a sample size smaller than 50 (18/33, 55%). Among the total of 7382 participants in all studies, the majority were patients diagnosed with breast cancer (1771/7382, 23.99%; Multimedia Appendix 3). A total of 27 studies (27/33, 82%) targeted patients on active anticancer treatment, and 13 studies (13/33, 39%) targeted patients with chemotherapy/endocrine therapy/immunotherapy (Tables 1 and 2). The majority of the studies had a quasi-experimental study design (25/33, 76%). The remaining studies used experimental (7/33, 21%), mixed methods (7/33, 21%), case control (1/33, 3%), and qualitative designs (1/33, 3%).

Table 1. Anticancer treatment type (N=33 studies).

Anticancer treatment types	Studies, n (%)
Surgery	6 (18)
Chemotherapy or hormonotherapy or immunotherapy	13 (39)
Radiation	7 (21)
All types mentioned	2 (6)
Unknown	5 (15)

Table 2. Anticancer treatment status (N=33 studies).

Anticancer treatment status	Studies, n (%)
On active treatment	27 (82)
Either on active treatment or survivors after treatment	2 (6)
Unknown	4 (12)

In addition to assessing patients' acceptance and use of home-based e-SRS in most studies, 1 study also reported caregivers' attitudes and preferences toward e-SRS [29]. One study reported that 8 out of 92 participants actually had their family caregivers who reported their symptoms for them, whereas there was no information provided regarding caregivers' acceptance and use in this study [30]. Study durations ranged from 1 month to 24 months, of which 7 studies (7/33, 21%) [31-37] followed up with participants for more than 6 months (Multimedia Appendix 2).

In total, 17 web-based e-SRS were reported in 22 studies (22/33, 66%) [27-29,31-49], including 2 studies that integrated web-based platforms with patient portals and electronic health records (EHRs) [41,47]. A total of 9 studies presented 7 mobile app–based e-SRS (9/33, 27%) [30,50-57], 1 study reported an interactive voice response system [58], and 1 study used only text messaging for symptom reporting [59]. The most commonly adopted symptom reporting instruments or questionnaires in home-based e-SRS were the National Cancer Institute-Common Terminology Criteria for Adverse Events or patient-reported outcome version of Common Terminology Criteria for Adverse Events (6/33, 19%) [27,28,30,38,46,48] and the European Organisation for Research and Treatment of Cancer Questionnaires (4/33, 12%) [28,31,45,49]. A total of 5 studies (5/33, 15%) required patients to report at least one of the several listed symptoms [40,42,54,57,58]. Most studies specified a reporting frequency, including daily (n=11), weekly (n=10), every other week (n=2), monthly (n=1), and less frequent than monthly (n=3), although a small number allowed patients to report their symptoms whenever they wanted (4/33, 12%) and 2 studies (2/33, 6%) did not report reporting frequencies (Multimedia Appendix 2).

e-SRS Acceptance

None of the 33 studies assessed patients' initial intention to use e-SRS before actual use. A total of 23 studies quantified patients' enrollment rates in e-SRS studies and reported a median rate of 68% (range, 22/25-18/18, 40%-100%) [27-30,32-34,39-42,44,46-49,51-53,55,56,58,59]. Mobile app–based systems showed lower enrollment rates than that of web-based systems (median 57% vs 71%). Among the 7 out of 23 studies that used mobile app–based systems [30,50-53,55,56], 4 studies (enrollment rates=40%-57%, 22/25-38/67) [50,51,53,56] showed that the most common reason for rejection was that patients did not have devices (eg, smartphones) or their devices were not compatible with the e-SRS platform (eg, iPhone or Android phone mismatched). A total of 2 studies using mobile app–based systems had relatively high enrollment rates (64/75, 85% and 66/107, 61.7%) because both studies provided mobile devices for participation [52,55].

A total of 7 studies assessed patients' technology acceptance after they used the systems [31,39-41,49,50,52]. Four of them reported that over 75% (56/75) of the patients stated, "I would continue to use it if asked." [31,39,40,52]. In addition, 5 studies reported that over 69% (59/85) stated, "I would recommend it to others." [40,41,49,50,52]. One study reported that 80%

(337/434) of the patients preferred e-SRS to the paper-and-pencil–based format in the future [31].

e-SRS Use

Patients' use of e-SRS was measured in various ways across the studies (Multimedia Appendix 4); one of the most common methods was to assess questionnaire response rates. However, the calculation of the response rate varied among studies. Several studies calculated the response rate using the number of patients who had ever reported their symptoms during certain time frames divided by the total number of all enrolled patients [29-32,36,40,42-44,46,48,50]. Overall, the mean percentage of the participants who had ever used e-SRS ranged from 70% (442/631) to 92% (45/49) across studies. A few other studies calculated the response rate using the number of submitted symptoms/forms divided by the total number of all expected forms. Using this method, the overall mean response rates across studies ranged from 45% (1596/3521) to 90% [29,35,37-39,42,49,52,53,55,56,58,59].

The log-on rate, that is, the frequency of accessing the system, was also adopted in some studies to measure the e-SRS use [30,33,39,45,46,53]. The log-on rate was calculated as the number of patients who logged on to systems divided by the number of all enrolled patients, which ranged from 33% (4/12) to 99% (99/100) across studies [30,34,45,46]. Some studies reported the average number of log-ons or the number of log-on days during the entire study period. For example, 1 study reported an average of 4 patient log-ons during a 30-day study period [42], another study reported an average of 17 log-ons over 34 weeks [36], and an average of 22 log-on days was reported during an average follow-up period of 12.70 months [53]. However, no study has reported the relationship between the log-on rates and the rates of actual symptom reporting.

Among the studies that assessed the change in e-SRS use over time [35,36,42,47,56], 2 longitudinal use patterns were identified. One pattern was the increased use from the beginning to nearly the midpoint of the study period (eg, initial 2 weeks of a 4-week study, 11-14 weeks of a 24-week study, or 16 weeks of a 34-week study), followed by a gradual decrease in use until the end of the study [35,36,42]. The second pattern was that the e-SRS use decreased over time throughout the study period [31,47,56]. For example, 85.1% (541/636) of patients used e-SRS within the first 6 months of one study, whereas the percentage decreased to 70% (442/631) at 9 months and 66.3% (414/624) at 15 months [31]. Overall, for the long-term use of e-SRS, the response rates ranged from 62% (40/64) to 85.1% (541/636) [31,33,35-37] and the log-on rates were 63.6% (103/162) to 77% (49/64) [33,44].

Facilitators/Barriers Associated With Home-Based e-SRS Acceptance and Use

Technology-Related Factors

The most commonly reported reasons for patients' reluctance to participate in e-SRS studies were the lack of access to compatible devices (eg, computers, smartphones, or tablets) [31,46,50,51,53,58]; lack of access to the internet [31,46]; or limited experience with computers, smartphones, or the internet [33,37,53,59]. A few studies excluded patients who did not have access to compatible devices or who did not have active email/patient portal accounts [36,39,40,42,43,51]. Only 1 study provided desktop computers [44], and 3 studies provided mobile devices to participants [52,55,57]. Patients with more technology experience had fewer technical issues and higher use of e-SRS [32,33,37,48,49]. Some patients could not use the systems owing to the failure of downloading apps [50] or incompatible operating systems with their devices [56].

e-SRS Features

Multiple studies reported patients' preferences for interactive system features, such as automatic reminders for symptom self-reporting [27,31,42,43,46,55,56] and health care providers' follow-up with self-reported symptoms [29,38,42,43,45,54,55]. System-generated self-management recommendations contributed to patients' high use and satisfaction [29,38,46,49,54,55]. Patients also favored the systems' features of (1) tracking symptoms over time [54,55], (2) bookmarking [31,34], (3) summarizing the symptom review [31], (4) having an icon- and image-based interface [30], (5) interacting with other patients [32], (6) reporting in free-text format [34], (7) connecting to EHR [41], and (8) interoperating with mobile devices [41].

Symptom Reporting Questionnaires

The quality of symptom questionnaires potentially affected patients' acceptance and use of e-SRS for symptom self-reporting [29,31,38,52,54,55]. Some patients complained about the overload or overlap of questions in the questionnaires [31,52,54] or questions that were difficult to understand [29,38,55]. Patients sometimes lost interest in using the system because the symptoms listed in the questionnaires were irrelevant to the symptoms they wanted to report, and the simple grading scale that requires patients to grade the presence or severity of certain symptoms was confusing [31,52,54].

Physical Health and Psychosocial Status

Health status of patients was associated with their acceptance and use of e-SRS. Patients often missed reporting of symptoms because of their illness, and about 7% of the missing data were due to the patients who were too ill to complete the questionnaires [36]. Some patients expressed their dislike or disinterest in participating in e-SRS studies because they felt that they were too tired (lack of energy) to engage in routine electronic symptom self-reporting [36,45,50,53,56,58,59]. Patients with brain tumors, such as glioma, struggled to use the technology because of the loss of their hand strength and poor memory [29]. Visual impairments in older people disrupted the use of electronic systems [34].

Qualitative interviews showed that patients' level of self-confidence and control in managing their health played an important role in their use of e-SRS [29,55], whereas some studies reported that increased symptom-related stress was associated with increased e-SRS use [32,34,42]. Some other studies indicated that patients might worry about the increased awareness of their symptoms through symptom tracking and reporting [33,46,52,53]. One study reported that patients were reluctant to use the system because they were afraid of being overly focused on their unpleasant symptoms, and the constant

detection of minor symptoms that they usually ignored might eventually make them mentally exhausted [29].

Home-Based Reporting

Patients were satisfied with their use of e-SRS at home because of (1) the convenience of flexible times and frequencies of reporting [29,31,34,43]; (2) timely symptom reporting, particularly for acute symptoms [31,43]; (3) benefits for patients who had concerns about language barriers [29] and who lived far from clinics [42]; and (4) reduction of clinic visit durations as clinicians have already been aware of their symptoms [39]. However, some patients might have concerns about the lack of face-to-face interactions with their providers by using e-SRS [29,38].

Demographic Factors

Patients who enrolled in the e-SRS studies had a mean age of 54 to 64 years, and patients who did not enroll had a mean age of 62.2 to 66 years [31,45,50,59]. Younger age [31,34,45,48,50,59], higher education level [36,38,48,49], White race [36,41,45], and male sex [31,36] were associated with higher acceptance and use of e-SRS. Evidence regarding the influence of employment status [49,56] and cancer staging [36,41] was mixed.

Discussion

Principal Findings

To the best of our knowledge, this study is the first to focus on acceptance and use of home-based e-SRS for symptom self-reporting by patients with cancer. This study also explored potential facilitators and barriers to e-SRS acceptance and use. Home-based e-SRS demonstrated the advantages of convenience, flexibility, and on-time symptom reporting. In addition, it enhanced patients' self-confidence in symptom control during/after their cancer treatment. However, considering the various participation rates and diverse reasons for nonuse of the systems reported, this study identified that the lack of technology compatibility was still a significant barrier to patients' adoption of home-based e-SRS. Although providing eHealth services or mobile devices to patients may help meet their needs for technology access, from the system design and development perspective, increasing the compatibility of e-SRS on multiple platforms seems to be a more potentially effective solution. Furthermore, system features, quality of symptom questionnaires, characteristics and health status of patients, and perceived benefits of using the systems were the important factors associated with acceptance and use of home-based e-SRS by patients with cancer. These findings are in line with literature reports that patient engagement in digital health interventions is associated with personal agency, motivation, and the quality of digital health interventions [22].

The review revealed that inconsistent approaches were used to assess e-SRS acceptance or use across studies, which might be because of different study purposes and designs. For example, the purpose of quasi-experimental pilot or feasibility studies was to investigate the feasibility of recruiting participants in e-SRS studies, in which enrollment rates served as an indirect assessment of patient acceptance of e-SRS [31]. Randomized controlled trials had aimed to evaluate the effects of the technology interventions on patient outcomes, in which participants' exposure to the technology systems was usually defined by a minimal threshold of access to the system or assessed based on patient outcomes of interest [49]. Inconsistent measurement and reporting of e-SRS acceptance and use across studies made the synthesis of findings challenging [19]. According to the widely adopted technology acceptance model, technology acceptance can be directly assessed by the person's statements regarding her/his behavioral intention to the use of technology [60]. The assessment of the actual use of technology is complex, as the expected outcomes of technology use vary by systems. Although log files are commonly used to measure how frequently users use technology systems, the log-in to the systems does not always accurately reflect users' actual use of the systems for expected outcomes. In the case of use of home-based e-SRS, users may log into the system, but they do not use the reporting function to report their symptoms. Overall, the measure of the actual use of technology should consider whether the users have performed the technology functionalities for expected outcomes.

Among all identified facilitators of and barriers to e-SRS acceptance and use, interactive features of e-SRS and the quality of symptom reporting questionnaires are considered as modifiable factors that can be purposefully modified and upgraded to meet users' needs, compared with nonmodifiable factors such as patients' demographic or clinical characteristics. In general, patients preferred regular reminders for their use of e-SRS and the feature of receiving feedback from either automated self-management advice or their health care team on their reported symptoms, which can be considered essential features in home-based e-SRS [19]. Although patients favored the integration of electronic symptom reporting with their electronic medical records, not many current home-based e-SRS have considered interoperability with clinical information systems [41,47]. Such limitations in system design and development should be addressed in future upgrades. In addition, a well-designed personalized questionnaire or a personalized way to deliver the questionnaire, for example, using a specific symptom-focused questionnaire or a specific type of treatment-focused questionnaire seems to encourage patients' use of e-SRS for voluntary symptom self-reporting.

This study suggested that clinicians' feedback on patients' symptom reporting through e-SRS was a facilitator of patient use of the system. The literature indicates that patients' use of e-SRS for symptom self-reporting provides opportunities for clinicians to understand patients' dynamic needs over time and facilitates productive interactions and interpersonal relationships between patients and clinicians [10,61]. Within the current outpatient oncology model, particularly with the increasing use of oral anticancer treatment at home, patients with cancer have fewer opportunities to directly interact with their health care professionals during office visits, share the concerns of their symptoms experienced, or receive sufficient information for symptom self-management [62]. The paradigm shift in cancer care delivery encourages the adoption of novel platforms for more effective patient-provider communication to support patient-centered care. Although the emerging home-based e-SRS

provides the opportunity to engage and empower patients and families in health communication and symptom self-management, the adoption of home-based e-SRS from the health system and health professional side remains unclear [21]. None of the reviewed studies reported clinicians' interactions with home-based e-SRS, that is, clinicians' acceptance and use of symptom information reported from the system. To fill this gap, the design and development of home-based e-SRS should consider providers' preferences for the way to interact with home-based e-SRS and health systems' expectations for the integration of home-based e-SRS into clinical workflows, to close the loop for optimal care delivery.

This study identified a minimal number of studies evaluating long-term e-SRS use (≥6 months) [31-37]. Despite the increasing recognition of the importance of patient symptom self-reporting throughout the cancer care trajectory, patients' long-term use of home-based e-SRS for symptom self-reporting tended to decline over time. The potential dynamic e-SRS use patterns identified in this study suggest that more studies are needed to increase our understanding of patients' long-term e-SRS use behaviors. Further exploration of factors associated with e-SRS long-term use trajectory patterns will contribute to the development of personalized support for patients' use of e-SRS for symptom self-reporting.

Patients' expectations and motivations for interacting with reporting systems may also change over time. For example, expectations for using e-SRS vary with patient health status. Interestingly, patients' health status could be either positively or negatively associated with their technology use behavior [20,63]. Patients with increased symptom distress might be motivated to continue tracking and reporting their symptoms to health care professionals. However, it was also possible that some patients decreased their use of e-SRS because they wanted to ignore the deterioration of their health status [36,46]. Further studies can explore the conditions and contexts that interfere with patients' use of e-SRS when their health status improves or declines.

Patients' feedback from postintervention interviews and surveys revealed that patients were more likely to continue using e-SRS after realizing that the systems were useful and convenient to use. Despite the low level of e-SRS acceptance (ie, low enrollment rates) before using the systems, patients' behavioral intention to use e-SRS improved after their actual use. We did not identify any study that provided patients with information or training before using a home-based e-SRS. Therefore, it is unknown whether patients' behavioral intentions, including perceived usefulness and ease of use, change after exposure to e-SRS. In future studies, it would be interesting to investigate how proactive and individualized training sessions reinforce patients' use of e-SRS for symptom self-reporting.

According to the UTAUT model, age, gender, and previous technology experiences potentially moderate the effects of determinants on the actual behavior of technology use [21]. Consistent with the UTAUT model, these personal factors were also identified in this study, especially age and previous technology experiences. These factors are not modifiable but may contribute to the development of targeted interventions and support for specific subgroups of the population. It is well known that family caregivers are commonly involved in medication and symptom self-management of patients with cancer, particularly for older adults with cancer [64]. This study indicated that family caregivers were sometimes the persons who actually used home-based e-SRS to report symptoms for their family members. It is also important to understand family caregivers' opinions of e-SRS use, as family caregivers' perceptions of symptom distress and symptom reporting may not always be congruent with those of patients with cancer [65]. However, there was a lack of research on cancer patient family caregivers' acceptance and use of e-SRS, which can be another important research topic in the future.

Patients appreciated that they had fewer time constraints in home-based reporting, and they reported their symptoms in real time without concerns of recalling their symptom experiences during their clinic visits. Furthermore, remote home-based symptom self-reporting could be especially beneficial for the underserved population who have geographic barriers to access health care or language problems for health care communications [29,42]. Of note, at the time of this study, the COVID-19 pandemic was pushing patients and clinicians to find new ways to work together. Perhaps now, more than ever, is the time to encourage patient adoption of e-SRS in cancer care in ways that can efficiently inform the conversation between clinicians and patients during a virtual telehealth clinical visit. The results of this study provide insight into how to engage patients in the use of e-SRS to facilitate telehealth care to improve health outcomes.

Limitations

This study has several limitations. First, many studies had small sample sizes and did not include diverse populations. Thus, there was a risk of selection bias. Second, the literature search was limited to studies published from January 2010 to March 2020 and to English language papers. Studies published before 2010 and studies published in non-English languages may contain useful information regarding patients' opinions on using e-SRS at home. Finally, this study considered participant enrollment rates in e-SRS studies as a surrogate measure of acceptance of e-SRS by patients with cancer. Such indirect measures might be less accurate, as some patients might refuse e-SRS studies for reasons that were not related to their acceptance to use e-SRS for symptom reporting. We focused on studies that reported the reasons for nonparticipation and extracted those that were potentially relevant to technology acceptance.

Conclusions

There is a growing interest in managing symptoms of patients with cancer remotely over time using electronic technology systems. Home-based e-SRS provides opportunities for patients with cancer to engage in symptom self-reporting from the initial cancer diagnosis, throughout treatment, and well into survivorship. It is important to evaluate patients' acceptance and use of e-SRS with standardized assessments so that the sustainability of the systems will be possible. Furthermore, understanding the facilitators and barriers of e-SRS regarding its acceptance and use in home settings will enhance the

dissemination of e-SRS in routine cancer care by patients with cancer. This study highlights the importance of assessing patients' accessibility to technology, physical and psychosocial status, and demographic factors for optimal symptom self-reporting. In addition, the design and development of interactive system features and personalized symptom reporting questionnaires should be considered to increase patient engagement. Future studies should explore long-term e-SRS use behavioral patterns of patients and develop personalized interventions to support symptom self-management and self-reporting for optimal health-related outcomes of patients with cancer.

Acknowledgments

The authors would like to acknowledge Ms Susan A Sung's contributions to language editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOC File , 71 KB - jmir_v23i3e24638_app1.doc]

Multimedia Appendix 2
Study characteristics.
[DOC File , 141 KB - jmir_v23i3e24638_app2.doc]

Multimedia Appendix 3
The number of participants by cancer type (n=7382 participants).
[PNG File , 192 KB - jmir_v23i3e24638_app3.png]

Multimedia Appendix 4
Measures of patient use of electronic symptom self-reporting system in the studies.
[DOCX File , 96 KB - jmir_v23i3e24638_app4.docx]

References

1. Basch E. The rationale for collecting patient-reported symptoms during routine chemotherapy. American Society of Clinical Oncology Educational Book 2014 May(34):161-165. [doi: 10.14694/edbook_am.2014.34.161]
2. Guidance for industry - patient-reported outcome measures: use in medical product development to support labeling claims. US Food and Drug Administration. 2009. URL: https://www.fda.gov/media/77832/download [accessed 2020-07-01]
3. Barkley R, Soobader M, Wang J, Blau S, Page RD. Reducing cancer costs through symptom management and triage pathways. JOP 2019 Feb;15(2):e91-e97. [doi: 10.1200/jop.18.00082]
4. Basch E. The missing voice of patients in drug-safety reporting. N Engl J Med 2010 Mar 11;362(10):865-869. [doi: 10.1056/nejmp0911494]
5. Basch E, Rogak LJ, Dueck AC. Methods for Implementing and Reporting Patient-reported Outcome (PRO) Measures of Symptomatic Adverse Events in Cancer Clinical Trials. Clin Ther 2016 Apr;38(4):821-830 [FREE Full text] [doi: 10.1016/j.clinthera.2016.03.011] [Medline: 27045992]
6. Ahlbrandt J, Lablans M, Glocker K, Stahl-Toyota S, Maier-Hein K, Maier-Hein L, et al. Modern information technology for cancer research: what's in IT for me? An overview of technologies and approaches. Oncology 2020 Nov 15;98(6):363-369 [FREE Full text] [doi: 10.1159/000493638] [Medline: 30439700]
7. Bennett AV, Jensen RE, Basch E. Electronic patient-reported outcome systems in oncology clinical practice. CA Cancer J Clin 2012 Jul;62(5):337-347 [FREE Full text] [doi: 10.3322/caac.21150] [Medline: 22811342]
8. Eid R, Haddad FG, Kourie HR, Kattan J. Electronic patient-reported outcomes: a revolutionary strategy in cancer care. Future Oncol 2017 Nov;13(27):2397-2399 [FREE Full text] [doi: 10.2217/fon-2017-0345] [Medline: 29121784]
9. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Controlled Clinical Trials 2003 Apr;24(2):182-199. [doi: 10.1016/S0197-2456(02)00320-3]
10. McCorkle R, Ercolano E, Lazenby M, Schulman-Green D, Schilling LS, Lorig K, et al. Self-management: Enabling and empowering patients living with cancer as a chronic illness. CA Cancer J Clin 2011 Jan;61(1):50-62 [FREE Full text] [doi: 10.3322/caac.20093] [Medline: 21205833]

11. Harris J, Cheevers K, Armes J. The emerging role of digital health in monitoring and supporting people living with cancer and the consequences of its treatments. Current Opinion in Supportive and Palliative Care 2018;12(3):268-275. [doi: 10.1097/spc.0000000000000362]
12. Carrasco S, Symes L. Patient use of electronic methods to self-report symptoms: an integrative literature review. ONF 2018 May 1;45(3):399-416. [doi: 10.1188/18.onf.399-416]
13. Greenlaw C, Brown-Welty S. A comparison of web-based and paper-based survey methods. Eval Rev 2009 Jul 15;33(5):464-480. [doi: 10.1177/0193841x09340214]
14. Holch P, Warrington L, Bamforth L, Keding A, Ziegler L, Absolom K, et al. Development of an integrated electronic platform for patient self-report and management of adverse events during cancer treatment. Ann Oncol 2017 Sep 1;28(9):2305-2311 [FREE Full text] [doi: 10.1093/annonc/mdx317] [Medline: 28911065]
15. Ruland CM, Holte HH, Røislien J, Heaven C, Hamilton GA, Kristiansen J, et al. Effects of a computer-supported interactive tailored patient assessment tool on patient care, symptom distress, and patients' need for symptom management support: a randomized clinical trial. J Am Med Inform Assoc 2010 Jul 01;17(4):403-410 [FREE Full text] [doi: 10.1136/jamia.2010.005660] [Medline: 20595307]
16. Seow H, Sussman J, Martelli-Reid L, Pond G, Bainbridge D. Do high symptom scores trigger clinical actions? An audit after implementing electronic symptom screening. JOP 2012 Nov;8(6):e142-e148. [doi: 10.1200/jop.2011.000525]
17. Coolbrandt A, Van den Heede K, Vanhove E, De Bom A, Milisen K, Wildiers H. Immediate versus delayed self-reporting of symptoms and side effects during chemotherapy: does timing matter? Eur J Oncol Nurs 2011 Apr;15(2):130-136. [doi: 10.1016/j.ejon.2010.06.010] [Medline: 20685164]
18. Jensen RE, Snyder CF, Abernethy AP, Basch E, Potosky AL, Roberts AC, et al. Review of electronic patient-reported outcomes systems used in cancer clinical care. J Oncol Pract 2014 Jul;10(4):e215-e222 [FREE Full text] [doi: 10.1200/JOP.2013.001067] [Medline: 24301843]
19. Warrington L, Absolom K, Conner M, Kellar I, Clayton B, Ayres M, et al. Electronic systems for patients to report and manage side effects of cancer treatment: systematic review. J Med Internet Res 2019 Jan 24;21(1):e10875 [FREE Full text] [doi: 10.2196/10875] [Medline: 30679145]
20. Kim J, Park H. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012 Oct;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
21. Venkatesh V, Morris M, Davis G. User acceptance of information technology: toward a unified view. MIS Quaterly 2003;27(3):425-478 [FREE Full text] [doi: 10.2307/30036540]
22. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
23. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (minors): development and validation of a new instrument. ANZ J Surg 2003 Sep;73(9):712-716. [doi: 10.1046/j.1445-2197.2003.02748.x] [Medline: 12956787]
24. CASP qualitative research checklist. Critical Appraisal Skills Programme. 2018. URL: https://casp-uk.net/wp-content/uploads/2018/01/CASP-Qualitative-Checklist-2018.pdf [accessed 2020-07-01]
25. Abraham NS, Byrne CJ, Young JM, Solomon MJ. Meta-analysis of well-designed nonrandomized comparative studies of surgical procedures is as good as randomized controlled trials. J Clin Epidemiol 2010 Mar;63(3):238-245. [doi: 10.1016/j.jclinepi.2009.04.005] [Medline: 19716267]
26. Njau B, Covin C, Lisasi E, Damian D, Mushi D, Boulle A, et al. A systematic review of qualitative evidence on factors enabling and deterring uptake of HIV self-testing in Africa. BMC Public Health 2019 Oct 15;19(1):1289 [FREE Full text] [doi: 10.1186/s12889-019-7685-1] [Medline: 31615461]
27. Hauth F, Bizu V, App R, Lautenbacher H, Tenev A, Bitzer M, et al. Electronic Patient-Reported Outcome Measures in Radiation Oncology: Initial Experience After Workflow ImplementationElectronic patient-reported outcome measures in radiation oncology: initial experience after workflow implementation. JMIR Mhealth Uhealth 2019 Jul 24;7(7):e12345 [FREE Full text] [doi: 10.2196/12345] [Medline: 31342906]
28. Peltola MK, Lehikoinen JS, Sippola LT, Saarilahti K, Mäkitie AA. A novel digital patient-reported outcome platform for head and neck oncology patients-a pilot study. Clin Med Insights Ear Nose Throat 2016;9:1-6 [FREE Full text] [doi: 10.4137/CMENT.S40219] [Medline: 27721662]
29. Boele FW, van Uden-Kraan CF, Hilverda K, Reijneveld JC, Cleijne W, Klein M, et al. Attitudes and preferences toward monitoring symptoms, distress, and quality of life in glioma patients and their informal caregivers. Support Care Cancer 2016 Jul 15;24(7):3011-3022 [FREE Full text] [doi: 10.1007/s00520-016-3112-7] [Medline: 26879825]
30. Bae WK, Kwon J, Lee HW, Lee S, Song E, Shim H, et al. Feasibility and accessibility of electronic patient-reported outcome measures using a smartphone during routine chemotherapy: a pilot study. Support Care Cancer 2018 Nov 7;26(11):3721-3728. [doi: 10.1007/s00520-018-4232-z] [Medline: 29732483]
31. Ashley L, Jones H, Thomas J, Newsham A, Downing A, Morris E, et al. Integrating patient reported outcomes with clinical cancer registry data: a feasibility study of the electronic Patient-Reported Outcomes From Cancer Survivors (ePOCS) system. J Med Internet Res 2013 Oct 25;15(10):e230 [FREE Full text] [doi: 10.2196/jmir.2764] [Medline: 24161667]

32. Børøsund E, Cvancarova M, Ekstedt M, Moore SM, Ruland CM. How user characteristics affect use patterns in web-based illness management support for patients with breast and prostate cancer. J Med Internet Res 2013 Mar 1;15(3):e34 [FREE Full text] [doi: 10.2196/jmir.2285] [Medline: 23454601]
33. Børøsund E, Cvancarova M, Moore SM, Ekstedt M, Ruland CM. Comparing effects in regular practice of e-communication and web-based self-management support among breast cancer patients: preliminary results from a randomized controlled trial. J Med Internet Res 2014 Dec 18;16(12):e295. [doi: 10.2196/jmir.3348]
34. Brochmann N, Zwisler A, Kjerholt M, Flachs EM, Hasselbalch HC, Andersen CL. A new internet-based tool for reporting and analysing patient-reported outcomes and the feasibility of repeated data collection from patients with myeloproliferative neoplasms. Qual Life Res 2016 Apr 11;25(4):835-846. [doi: 10.1007/s11136-015-1125-1] [Medline: 26358064]
35. Iivanainen S, Alanko T, Peltola K, Konkola T, Ekström J, Virtanen H, et al. ePROs in the follow-up of cancer patients treated with immune checkpoint inhibitors: a retrospective study. J Cancer Res Clin Oncol 2019 Jan 21;145(3):765-774. [doi: 10.1007/s00432-018-02835-6]
36. Judson TJ, Bennett AV, Rogak LJ, Sit L, Barz A, Kris MG, et al. Feasibility of long-term patient self-reporting of toxicities from home via the Internet during routine chemotherapy. J Clin Oncol 2013 Jul 10;31(20):2580-2585 [FREE Full text] [doi: 10.1200/JCO.2012.47.6804] [Medline: 23733753]
37. Lucas SM, Kim T, Ghani KR, Miller DC, Linsell S, Starr J, Michigan Urological Surgery Improvement Collaborative. Establishment of a web-based system for collection of patient-reported outcomes after radical prostatectomy in a statewide quality improvement collaborative. Urology 2017 Sep;107:96-102. [doi: 10.1016/j.urology.2017.04.058] [Medline: 28652160]
38. Cowan R, Suidan R, Andikyan V, Rezk Y, Einstein H, Chang K, et al. Electronic patient-reported outcomes from home in patients recovering from major gynecologic cancer surgery: A prospective study measuring symptoms and health-related quality of life. Gynecologic Oncology 2016 Jun;141:175. [doi: 10.1016/j.ygyno.2016.04.455]
39. Denis F, Viger L, Charron A, Voog E, Dupuis O, Pointreau Y, et al. Detection of lung cancer relapse using self-reported symptoms transmitted via an internet web-application: pilot study of the sentinel follow-up. Support Care Cancer 2014 Jun;22(6):1467-1473. [doi: 10.1007/s00520-013-2111-1] [Medline: 24414998]
40. Denis F, Voog E, Pointreau Y, Bourgeois H, Seegers V, Le Du K. Prospective study of a web-mediated management of febrile neutropenia related to chemotherapy (Bioconnect). Support Care Cancer 2019 Jun 10;27(6):2189-2194. [doi: 10.1007/s00520-018-4505-6] [Medline: 30306327]
41. Garcia SF, Wortman K, Cella D, Wagner LI, Bass M, Kircher S, et al. Implementing electronic health record-integrated screening of patient-reported symptoms and supportive care needs in a comprehensive cancer center. Cancer 2019 Nov 15;125(22):4059-4068 [FREE Full text] [doi: 10.1002/cncr.32172] [Medline: 31373682]
42. Graetz I, Anderson JN, McKillop CN, Stepanski EJ, Paladino AJ, Tillmanns TD. Use of a web-based app to improve postoperative outcomes for patients receiving gynecological oncology care: A randomized controlled feasibility trial. Gynecol Oncol 2018 Aug;150(2):311-317. [doi: 10.1016/j.ygyno.2018.06.007] [Medline: 29903391]
43. Graetz I, McKillop CN, Stepanski E, Vidal GA, Anderson JN, Schwartzberg LS. Use of a web-based app to improve breast cancer symptom management and adherence for aromatase inhibitors: a randomized controlled feasibility trial. J Cancer Surviv 2018 Aug 28;12(4):431-440 [FREE Full text] [doi: 10.1007/s11764-018-0682-z] [Medline: 29492753]
44. Innominato P, Komarzynski S, Karaboué A, Ulusakarya A, Bouchahda M, Haydar M, et al. Home-Based e-Health Platform for Multidimensional Telemonitoring of Symptoms, Body Weight, Sleep, and Circadian Activity: Relevance for Chronomodulated Administration of Irinotecan, Fluorouracil-Leucovorin, and Oxaliplatin at Home-Results From a Pilot Study. JCO Clin Cancer Inform 2018 Dec;2:1-15 [FREE Full text] [doi: 10.1200/CCI.17.00125] [Medline: 30652550]
45. Sierra H, Cordova M, Chen CJ, Rajadhyaksha M. Confocal imaging-guided laser ablation of basal cell carcinomas: an ex vivo study. J Invest Dermatol 2015 Feb;135(2):612-615 [FREE Full text] [doi: 10.1038/jid.2014.371] [Medline: 25178106]
46. Warrington L, Absolom K, Holch P, Gibson A, Clayton B, Velikova G. Online tool for monitoring adverse events in patients with cancer during treatment (eRAPID): field testing in a clinical setting. BMJ Open 2019 Feb 1;9(1):e025185 [FREE Full text] [doi: 10.1136/bmjopen-2018-025185] [Medline: 30782751]
47. Zylla DM, Gilmore GE, Steele GL, Eklund JP, Wood CM, Stover AM, et al. Collection of electronic patient-reported symptoms in patients with advanced cancer using Epic MyChart surveys. Support Care Cancer 2020 Jul 8;28(7):3153-3163. [doi: 10.1007/s00520-019-05109-0] [Medline: 31701269]
48. Andikyan V, Rezk Y, Einstein MH, Gualtiere G, Leitao MM, Sonoda Y, et al. A prospective study of the feasibility and acceptability of a Web-based, electronic patient-reported outcome system in assessing patient recovery after major gynecologic cancer surgery. Gynecol Oncol 2012 Nov;127(2):273-277 [FREE Full text] [doi: 10.1016/j.ygyno.2012.07.124] [Medline: 22871467]
49. Berry DL, Blonquist TM, Patel RA, Halpenny B, McReynolds J. Exposure to a patient-centered, web-based intervention for managing cancer symptom and quality of life issues: impact on symptom distress. J Med Internet Res 2015 Jun 3;17(6):e136 [FREE Full text] [doi: 10.2196/jmir.4190] [Medline: 26041682]
50. Belarmino A, Walsh R, Alshak M, Patel N, Wu R, Hu JC. Feasibility of a mobile health application to monitor recovery and patient-reported outcomes after robot-assisted radical prostatectomy. Eur Urol Oncol 2019 Jul;2(4):425-428. [doi: 10.1016/j.euo.2018.08.016] [Medline: 31277778]

51. Falchook AD, Tracton G, Stravers L, Fleming ME, Snavely AC, Noe JF, et al. Use of mobile device technology to continuously collect patient-reported symptoms during radiation therapy for head and neck cancer: A prospective feasibility study. Adv Radiat Oncol 2016 Apr;1(2):115-121 [FREE Full text] [doi: 10.1016/j.adro.2016.02.001] [Medline: 28740878]
52. Furlong E, Darley A, Fox P, Buick A, Kotronoulas G, Miller M, et al. Adaptation and implementation of a mobile phone-based remote symptom monitoring system for people with cancer in Europe. JMIR Cancer 2019 Mar 14;5(1):e10813 [FREE Full text] [doi: 10.2196/10813] [Medline: 30869641]
53. Greer JA, Jacobs JM, Pensak N, Nisotel LE, Fishbein JN, MacDonald JJ, et al. Randomized trial of a smartphone mobile app to improve symptoms and adherence to oral therapy for cancer. J Natl Compr Canc Netw 2020 Feb;18(2):133-141. [doi: 10.6004/jnccn.2019.7354] [Medline: 32023526]
54. Gustavell T, Langius-Eklöf A, Wengström Y, Segersvärd R, Sundberg K. Development and feasibility of an interactive smartphone app for early assessment and management of symptoms following pancreaticoduodenectomy. Cancer Nurs 2019;42(3):E1-E10. [doi: 10.1097/ncc.0000000000000584]
55. Langius-Eklöf A, Christiansen M, Lindström V, Blomberg K, Hälleberg Nyman M, Wengström Y, et al. Adherence to report and patient perception of an interactive app for managing symptoms during radiotherapy for prostate cancer: descriptive study of logged and interview data. JMIR Cancer 2017 Oct 31;3(2):e18 [FREE Full text] [doi: 10.2196/cancer.7599] [Medline: 29089290]
56. Min YH, Lee JW, Shin Y, Jo M, Sohn G, Lee J, et al. Daily collection of self-reporting sleep disturbance data via a smartphone app in breast cancer patients receiving chemotherapy: a feasibility study. J Med Internet Res 2014 May;16(5):e135 [FREE Full text] [doi: 10.2196/jmir.3421] [Medline: 24860070]
57. Sundberg K, Eklöf AL, Blomberg K, Isaksson A, Wengström Y. Feasibility of an interactive ICT-platform for early assessment and management of patient-reported symptoms during radiotherapy for prostate cancer. Eur J Oncol Nurs 2015 Oct;19(5):523-528. [doi: 10.1016/j.ejon.2015.02.013] [Medline: 25813529]
58. Mooney KH, Beck SL, Wong B, Dunson W, Wujcik D, Whisenant M, et al. Automated home monitoring and management of patient-reported symptoms during chemotherapy: results of the symptom care at home RCT. Cancer Med 2017 Mar;6(3):537-546 [FREE Full text] [doi: 10.1002/cam4.1002] [Medline: 28135050]
59. Mougalian SS, Epstein LN, Jhaveri AP, Han G, Abu-Khalaf M, Hofstatter EW, et al. Bidirectional text messaging to monitor endocrine therapy adherence and patient-reported outcomes in breast cancer. JCO Clinical Cancer Informatics 2017 Nov(1):1-10. [doi: 10.1200/cci.17.00015]
60. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
61. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Q 2001 Sep;25(3):351. [doi: 10.2307/3250921]
62. Boons CC, Timmers L, van Schoor NM, Swart EL, Hendrikse NH, Janssen JJ, et al. Patient satisfaction with information on oral anticancer agent use. Cancer Med 2018 Jan 23;7(1):219-228 [FREE Full text] [doi: 10.1002/cam4.1239] [Medline: 29168352]
63. Jiang Y, Sereika SM, Dabbs AD, Handler SM, Schlenk EA. Acceptance and use of mobile technology for health self-monitoring in lung transplant recipients during the first year post-transplantation. Appl Clin Inform 2016;7(2):430-445 [FREE Full text] [doi: 10.4338/ACI-2015-12-RA-0170] [Medline: 27437052]
64. Jayani R, Hurria A. Caregivers of older adults with cancer. Semin Oncol Nurs 2012 Nov;28(4):221-225 [FREE Full text] [doi: 10.1016/j.soncn.2012.09.004] [Medline: 23107179]
65. Broberger E, Tishelman C, von Essen L. Discrepancies and similarities in how patients with lung cancer and their professional and family caregivers assess symptom occurrence and symptom distress. J Pain Symptom Manage 2005 Jun;29(6):572-583 [FREE Full text] [doi: 10.1016/j.jpainsymman.2004.11.006] [Medline: 15963866]

Abbreviations

CASP: Critical Appraisal Skills Programme
EHR: electronic health record
e-SRS: electronic symptom self-reporting system
MeSH: Medical Subject Headings
MINORS: Methodological Index for Nonrandomized Studies
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by G Eysenbach; submitted 28.09.20; peer-reviewed by Z Wang, A Darley, K Sundberg; comments to author 23.11.20; revised version received 18.01.21; accepted 25.01.21; published 12.03.21.

Please cite as:
Cho Y, Zhang H, Harris MR, Gong Y, Smith EL, Jiang Y
Acceptance and Use of Home-Based Electronic Symptom Self-Reporting Systems in Patients With Cancer: Systematic Review
J Med Internet Res 2021;23(3):e24638
URL: https://www.jmir.org/2021/3/e24638
doi:10.2196/24638
PMID:33709929

Review

Impact of Web-Based Sharing and Viewing of Self-Harm–Related Videos and Photographs on Young People: Systematic Review

Amanda Marchant[1], MSc; Keith Hawton[2], FMedSci, DSc; Lauren Burns[1], MSc; Anne Stewart[2], FRCPsych; Ann John[1], FFPH

[1]Swansea University Medical School, Swansea, United Kingdom

[2]Centre for Suicide Research, University Department of Psychiatry, Oxford, United Kingdom

Corresponding Author:
Ann John, FFPH
Swansea University Medical School
Singleton Park
Swansea, SA2 8PP
United Kingdom
Phone: 44 1792602568
Email: A.John@swansea.ac.uk

Abstract

Background: Given recent moves to remove or blur self-harm imagery or content on the web, it is important to understand the impact of posting, viewing, and reposting self-harm images on young people.

Objective: The aim of this study is to systematically review research related to the emotional and behavioral impact on children and young people who view or share web-based self-harm–related videos or images.

Methods: We searched databases (including Embase, PsychINFO, and MEDLINE) from January 1991 to February 2019. Search terms were categorized into internet use, images nonspecific and specific to the internet, and self-harm and suicide. Stepwise screening against specified criteria and data extraction were completed by two independent reviewers. Eligible articles were quality assessed, and a narrative synthesis was conducted.

Results: A total of 19 independent studies (20 articles) were included. Of these, 4 studies focused on images, 10 (11 articles) on videos, and 5 on both. There were 4 quantitative, 9 qualitative, and 7 mixed methods articles. In total, 11 articles were rated as high quality. There has been an increase in graphic self-harm imagery over time. Potentially harmful content congregated on platforms with little moderation, anonymity, and easy search functions for images. A range of reactions and intentions were reported in relation to posting or viewing images of self-harm: from empathy, a sense of solidarity, and the use of images to give or receive help to potentially harmful ones suggesting new methods, normalization, and exacerbation of self-harm. Viewing images as an alternative to self-harm or a creative outlet were regarded in 2 studies as positive impacts. Reactions of anger, hostility, and ambivalence have been reported. There was some evidence of the role of imitation and reinforcement, driven partly by the number of comments and wound severity, but this was not supported by time series analyses.

Conclusions: Although the results of this review support concern related to safety and exacerbation of self-harm through viewing images of self-harm, there may be potential for positive impacts in some of those exposed. Future research should evaluate the effectiveness and potential harms of current posting restrictions, incorporate user perspectives, and develop recovery-oriented content. Clinicians assessing distressed young people should ask about internet use, including access to self-harm images, as part of their assessment.

(J Med Internet Res 2021;23(3):e18048) doi:10.2196/18048

KEYWORDS

self-harm; suicide; social media; internet; systematic review

Introduction

Background

Young people who self-harm engage in more web-based activities than others of a similar age [1]. There is a large presence of self-harm–related materials [2], such as blogs [3,4], large social media groups [5], thousands of highly viewed videos [6], websites [7], and dedicated online communities [8]. The nature of internet use is constantly evolving. Platforms such as Instagram and Tumblr have increased in popularity for self-harm communities, partly because they are image based [9]. A total of 44.00% (528/1200) of surveyed adolescents reported Instagram to be an important part of their daily lives [10], and internet searches for suicide are increasingly returning graphic imagery [11].

Mental imagery plays a role in determining future behaviors [12]. It is thought to be more emotionally evocative than verbal thoughts, with stronger links to affect [13]. Individuals who experience suicidal ideation often report *flash-forward* detailed imagery, similar to flashback imagery in posttraumatic stress disorder, related to suicidal acts [14,15]. Most youths who self-injured reported mental imagery when the urge to self-injure was strong [16,17]. Emotion-laden mental imagery of self-harm may play a role in motivating behaviors toward or avoiding self-harm [17].

In addition to the role of mental imagery, a growing body of literature has explored the role of exposure to self-harm images through the media [8,18,19]. Although self-harm is a largely hidden behavior, images of self-harm are commonly shared on the internet. This is often done anonymously, with individuals continuing to hide their self-harm in the offline world [20]. Participants have reported reduced loneliness and the use of images to curb self-harm urges on the one hand and reinforcement and triggering of self-harm on the other [8]. Previous systematic reviews regarding internet use and self-harm, although not focused on images, similarly suggest both positive and negative impacts of images and videos [18,19]. The growing popularity of image-based platforms and high-profile stories of the risks has stimulated an exponential increase in research related to web-based self-harm imagery. Given recent moves to remove or blur self-harm imagery or content posted on platforms such as Instagram in response to concerns of bereaved parents [21], it is important to understand the impact of posting or viewing self-harm images on young people.

Aims

This study aims to systematically review research related to the viewing or sharing of web-based self-harm or suicidal behavior–related videos or images in children and young people to explore the impact on emotions and behaviors of viewing or reposting images or videos of self-harm, the impact of posting images or videos of self-harm by individuals who self-harm, and whether certain aspects or types of images or videos impact outcomes.

Methods

Previous Reviews

Web-based content related to video and images of self-harm has been included as part of previous systematic reviews by our research group [18,19], but it was not our main focus. We adapted the search strategy used to specifically identify research on this topic. An electronic literature search was conducted from January 01, 1991, (the year the internet was made publicly available) to February 20, 2019, to ensure a comprehensive overview of the existing literature.

Search Strategy

The core databases CINAHL, Embase (excluding MEDLINE journals), PsychINFO, SCOPUS, and MEDLINE were searched alongside topic-specific websites (Campbell, Centre for Mental Health, Department of Health, Mental Health Foundation, Department of Health, Social Services and Public Safety for Northern Ireland, National Health Service (NHS) Scotland, and the Royal College of Psychiatrists) and meta-search engines (Google and Google Scholar). The search terms were grouped into 4 categories:

1. Internet use: Free text “Aol” or “Askfm” or “Bebo” or “blog*” or “chat?room*” or “cyber*” or “discussion forum” or “e?communi*” or “e?material$” or “Facebook” or “googl*” or “hashtag$” or “instant messag*” or “internet” or “live chat*” or “live journal$” or “MSN” or “Myspace” or “on?line” or “online” or “podcast*” or “social network*” or “spam*” or “tweet*” or “Twitter” or “troll” or “virtual*” or “web” or “whatsapp”. *MeSH terms* “Internet,” “Social media,” “Social networking”
2. Images nonspecific to internet: Free text. “imag*” or “galler*” or “photo*” or “picture$” or “video*”. *MeSH terms* “Video recording”
3. Images specific to the internet: Free text. “meme” or “Pinterest” or “Tumblr” or “vine” or “vlog*” or “YouTube” or “snap*” or “gif$” or “selfie$” or “Flikr” or “camera” or “filter” or “reddit” or “Instagram” or “tik tok”
4. Self-harm and suicide: Free text. “Automutilation” or “Distress*” or “emotion*” or “NSSI” or “SIB” or “suicid*” or ((oneself or myself or self) adj2(cut* or harm* or hurt* or kill or injur* or mutilat*)). *MeSH terms* ‘Self-injurious behaviour’, ‘Stress, Psychological’

The terms were combined as follows (internet AND images nonspecific to internet terms) OR (images specific to the internet). The self-harm and suicide groups were combined with the results from the above search.

Selection Criteria

Articles were included if they examined web-based viewing or sharing of images or videos related to self-harm or suicidal behavior or, videos and images viewed or shared by individuals who experienced suicidal ideation, suicidal behavior, or self-harm. Self-harm refers to any intentional act of self-injury or poisoning, independent of motivation or suicidal intent [22]. This definition is very general, as the motivation or degree of suicidal intent is difficult to assess and may vary between individuals and over time. Articles were required to include

primary empirical data and be published in peer-reviewed journals. The results were not restricted by the location. Only English language articles were included. Review articles, single case studies, editorials, conference abstracts, or other grey literature were not included. The reference lists of all review articles were manually screened for potential eligible articles. Participants had to be aged 24 years or under or the study population had to have an average age of 24 years and under. If age was not specified, participants had to be described as children, adolescents, young people, or young adults. Where articles examined more than 1 age group, only data for the age group fitting these criteria were analyzed.

Two independent reviewers (AM and LB) manually screened titles. Any disagreements were resolved by consensus or discussed with a third expert reviewer (AJ). Duplicates were removed. Titles that had no relevance and gray literature were excluded from the title screen. A record was kept of all discarded articles, including the reason for exclusion. The remaining titles with abstracts were screened for eligibility (AM and LB). Full-text articles were obtained where suitability could not be determined based on the title and abstract.

Data Analysis

A previously developed data extraction sheet [19] was used to record the findings (Multimedia Appendix 1). Additional fields were added to examine features specific to images and videos (eg, platform moderation, trigger warnings, the nature of the images or videos, and the impact on the viewer). Two reviewers (AM and LB) independently extracted the data from each article. Any inconsistencies in data extraction and quality scores were clarified by consensus or through discussion with a third expert reviewer (AJ).

The heterogeneous and largely qualitative nature of the studies precluded any meaningful combination of results through meta-analysis. Therefore, narrative synthesis was employed. On the basis of published guidance [23], this narrative synthesis examined a number of key aspects of each article. Comparisons were made regarding the way in which self-harm imagery was examined across studies and internet platforms. The influence of heterogeneity was explored further, including differences in populations studied, differences between various platforms, measures employed, and outcomes studied.

Articles were amalgamated and grouped into categories. These categories were inductively generated following initial reading and data extraction of articles and were cross-checked by members of the study team (AM, LB, and AJ). Positive influences of viewing or sharing of images and videos were defined as a perceived reduction in psychological distress, reduced suicidal ideation or self-harm, and advice on how to seek help and encouragement to do so. Negative influences were defined as results indicating increased psychological distress, increased frequency or severity of self-harm or suicidal ideation, provision of information on methods of self-harm or suicide, and self-harm behaviors being encouraged [19]. Positive and negative influences were examined in relation to the population studied, internet platforms (eg, Instagram and Tumblr), and specific features of images and videos.

The quality of the included articles was assessed according to the Critical Appraisal Skills Programme (CASP) [24], as described previously [19]. The CASP tool assesses various aspects of study design, including sources of bias, data collection, clarity of results, and appropriateness of conclusions. CASP does not specifically recommend any scoring or grading system. We adopted a scoring system used previously [25] (≤50% of criteria low quality, 50%-74% moderate, ≥75% high quality, and no criteria weighting applied [26]). Quality assessment was conducted by two independent reviewers (AM and LB) for each article.

A protocol for the systematic review was registered with PROSPERO (international prospective register of systematic reviews) [27].

Results

Screening Process

Figure 1 shows the results of the search strategy and screening process. A total of 19 independent studies (20 articles) were included. Of which, 4 studies focused on images, 10 studies (11 articles) on videos, and 5 on both images and videos. Two articles were based on a single data set of 100 videos [6,28]. A summary of the included articles, quality scores, by category of study is presented in Tables 1-3. Studies came from across the world (United States, n=7; Canada, n=4; United Kingdom, n=2; Germany, n=2; Australia, n=1; and multiple countries, n=3). There were 4 quantitative, 9 qualitative, and 7 mixed methods articles. In total, 11 articles were rated as high quality, 7 as moderate, and 2 as low. Tables 1-3 provide summaries of the included studies.

Figure 1. Results of the search strategy and screening process.

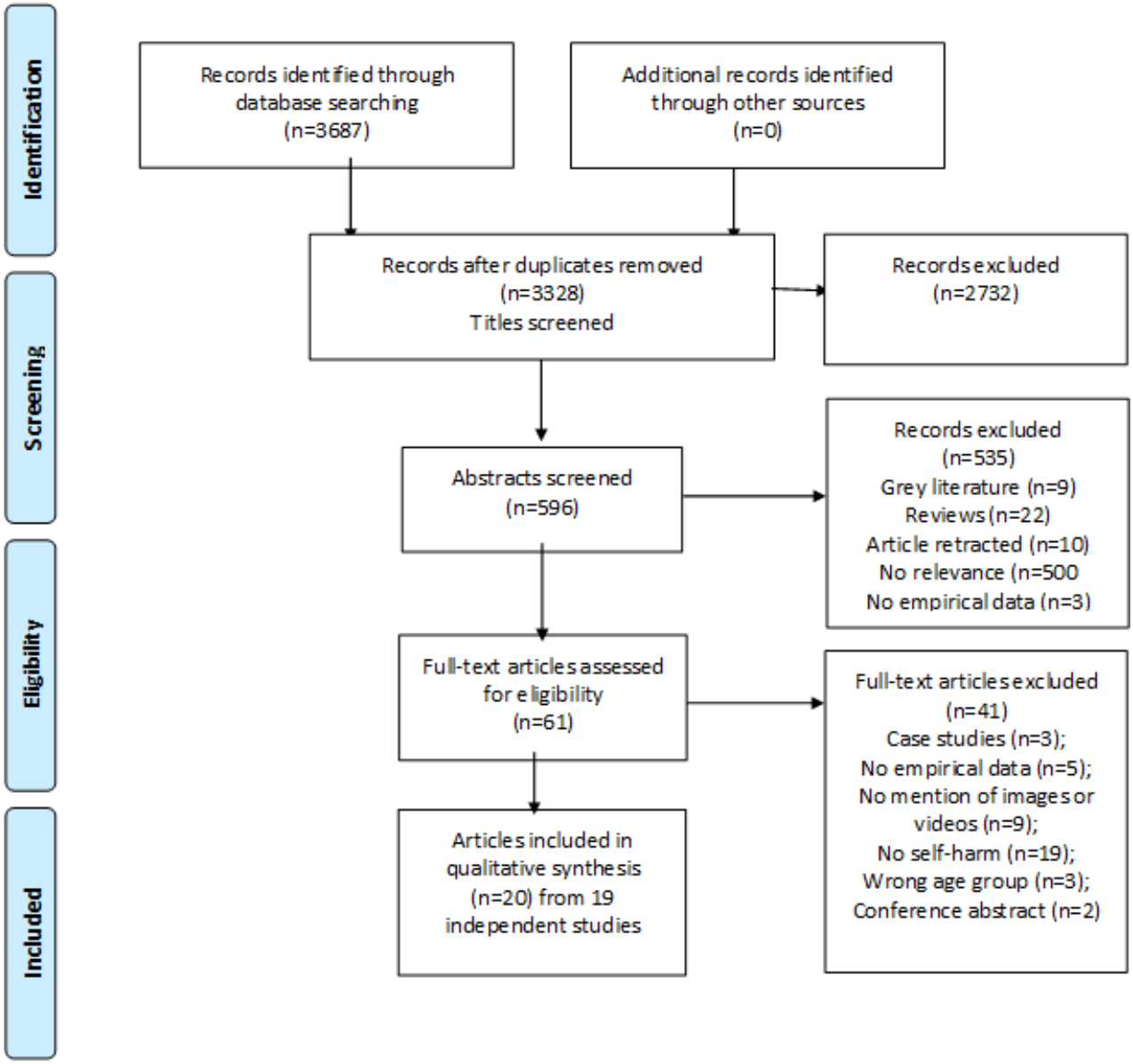


Table 1. Summary of included studies analyzing images or videos related to self-harm or suicidal behaviors.

Category (Author, year, country, quality score)	Analysis of images or videos related to self-harm or suicidal behaviors		
	Aims and objectives	Participants or sample	Results
Avery, 2015 [29], United States, moderate.	To understand the characteristics of the fire challenge phenomenon in an effort to develop preventative and public safety measures	50 videos depicting the fire challenge; 90% (45/50) male; 64% (32/50) African American	50 videos were selected. Of these, 13 videos included postburn footage. Superficial and partial thickness burns were sustained on the torso (10/13, 77%), face (4/13, 31%), and extremities (2/13, 15%). Full thickness burns were seen in 15% (2/13) videos. Young African American males were overrepresented. The authors suggested that these should be targeted in prevention and intervention.
Basch, 2017 [30], United States, moderate.	To describe the extent to which content related to bullying is present on YouTube with respect to source, content, number of views, length, and year uploaded	The top 100 videos related to bullying on YouTube with the greatest number of total views were identified.	The most common content was describing or depicting violence (n=89). Over half addressed getting help (n=56). A total of 38 out of 100 videos mention suicide or thoughts of suicide. No analysis was conducted of the nature of discussions around suicide or their potential impact.
Brown, 2018 [31], Germany, high.	To analyze pictures directly depicting self-harm wounds on Instagram. Pictures, comments, and user accounts were examined. The aims were to systematically describe the extent of self-harm on Instagram in a German-speaking population and to describe web-based content of German-speaking users	Posters of self-harm pictures on Instagram (n=1152); majority anonymous; 91.0% (745/819) of identifiable profiles, female	Most pictures depicted wounds caused by cutting on arms or legs and were rated as mild or moderate injuries. Pictures with increasing wound grades and those depicting multiple methods of self-harm generated elevated amounts of comments. Most comments were neutral or empathic, with some offering help. Few comments were hostile. Pictures were mainly posted in the evening with a small peak in the early morning and on Sundays.
Cavazos-Rehg, 2016 [32], United States, high.	To gain a better understanding of the depression, self-harm, and suicidal content that is being shared on Tumblr	Tumblr accounts; of the 8 that provided demographic information 75% (6/8) were female.	A total of 17 Tumblr accounts posted a median number of 185 posts. Content was engaged with 1,677,362 times. Of the 3360 randomly selected posts, 81.52% (2739/3360) were related to depression, suicide, or self-harm. Common themes were self-loathing (412/2739, 15.04%), loneliness or feeling unloved (405/2739, 14.79%), self-harm (407/2739, 14.86%), and suicide (372/2739, 13.58%). Findings signal a need for suicide prevention efforts to intervene on Tumblr and use this platform in a strategic way.
Duggan, 2012 [2], Canada, low.	To simultaneously examine the scope and nature of self-harm content across informational or interactive websites, social networking websites, and YouTube to provide mental health practitioners with a multifaceted description of web-based content related to self-harm	Demographics only available for uploaders of YouTube videos; 80% (4/5) were female.	Peer-driven, informal websites have a variety of triggering content and are accessed more often than professionally driven sites. Self-harm is strongly represented among social networking websites and YouTube, evidenced by large group memberships and large video view counts.

Category (Author, year, country, quality score)	Analysis of images or videos related to self-harm or suicidal behaviors		
	Aims and objectives	Participants or sample	Results
Grzanka, 2014 [33], United States, low.	To investigate the mass media "It Gets Better" campaign responding to a perceived increase in suicides among gay youth in the United States since September 2010; critical discourse analysis of a sample of videos posted as part of campaign	Content of 30 videos aimed at LGBT+[a] youth was analyzed.	Activists, celebrities, and the general public created and uploaded videos as part of the campaign, often telling their own stories of overcoming difficulties (eg, bullying). A thematic analysis of a subset of these videos revealed common themes of directives (eg, "don't give up") and testimonies of how the posters had overcome their own difficulties. The effectiveness of the campaign was not analyzed. Results revealed a neoliberal framing that placed the burden of a better life onto the emotional lives of LGBT+ youth, who were instructed to endure their pain in the interest of inevitable happiness.
Hilton, 2017 [34], United Kingdom, high.	To report the findings from a unique analysis of naturally occurring data regarding self-harm behavior generated through the global social media site, Twitter	Twitter users n=317; no information regarding gender	A total of 5 themes were identified: (1) celebrity influence, (2) self-harm is not a joke, (3) support for and from others, (4) eating disorders and self-harm, and (5) videos and personal stories. More recovery-oriented content than graphic imagery was found on this platform. No formal comparison with other platforms is made.
Lewis, 2012[b] [28], Canada, high.	To examine viewers' comments and responses to self-harm YouTube videos to determine the potential risks and benefits of the videos	Uploaders of 100 self-harm videos on YouTube; 95.0% (95/100) female	Most frequent comments were self-disclosure (334/869, 38.4%) and individuals sharing their own self-harm experiences. This was followed by feedback for the video uploader, including admiration of video quality (191/869, 22.0%) or message (148/869, 17.0%), and admiration for uploader (134/869, 15.4%), or encouragement for uploader (97/869, 11.2%). Many of those self-disclosing did not mention recovery (374/869, 43.0%) and indicated they were still self-injuring (295/869, 34.0%).
Lewis, 2011[b] [6], Canada, moderate.	To explore the accessibility and scope of web-based self-harm videos	Uploaders of 100 self-harm videos on YouTube; 95.0% (95/100) female	The top 100 videos analyzed were viewed over 2 million times, and 80.0% (80/100) were accessible to a general audience. Viewers rated the videos positively and selected videos as a favorite over 12,000 times. The tone of the videos was largely factual or educational (53/100, 53.0%) or melancholic (eg, hopeless statements, depictions of sadness or crying; 51.0% [51/100]). Explicit imagery of self-injury was common. A total of 90.0% (90/100) of noncharacter videos had self-harm photos, whereas 28.0% (28/100) of character videos had inaction self-harm. For both, cutting was the most common method. Many videos (58/100, 58.0%) did not warn about this content. Content was often creative and frequently contained graphic imagery.

Category (Author, year, country, quality score)	Analysis of images or videos related to self-harm or suicidal behaviors		
	Aims and objectives	Participants or sample	Results
Lewis, 2015 [35], Canada, high.	To examine the nature of self-harm first aid tip videos shared through YouTube	40 uploaders of videos to YouTube; 83% (33/40) female	A total of 40 self-harm YouTube videos were content-analyzed. Videos were viewed 157,571 times and typically favorably rated. Most had a neutral purpose and neither encouraged nor discouraged self-harm. Messages encouraging self-harm help seeking were scant. Medical help seeking was not commonly encouraged 55% (22/40) of videos related to self-cutting and 28% (11/40), related to self-burning recommended seeking medical help, with several videos providing "safe" self-harm instructions.
Moreno, 2016 [36], United States, high.	To evaluate the meaning, popularity, and content advisory warnings relating to ambiguous self-harm hashtags on Instagram by (1) presenting current data on ambiguous hashtags that may be commonly related to self-harm; (2) testing a process to investigate ambiguous self-harm terms; (3) evaluating the popularity of self-harm-related hashtags at 2 time points; and (4) assessing the precision of Instagram's warning labels for concerning content	Posters of ambiguous self-harm hashtags on Instagram; 193 unique usernames in a sample of 201 posts; no information on gender	Sample of 201 Instagram posts led to identification of 10 ambiguous self-harm hashtags. Results demonstrated a popular image that described the broader community of self-harm and mental illness called #mysecretfamily that had approximately 900,000 search results at time 1 and 1.5 million at time 2. Only one-third of relevant hashtags generated content advisory warnings.
Poonai, 2018 [37], Canada, high.	To use interrupted time series analysis to examine whether the release of Amanda Todd's YouTube video following her death announcement in October 2012 was associated with an increase in average monthly emergency department visit rates for suicide-related diagnoses in Ontario children aged 11-17 years	Population-based time series analysis using a national database. Ontario patients aged 11-17 years based on a sample of 4,775,658 emergency department visits; 48.87% (2,333,822/4,775,658) female	There was a significant increase in the monthly emergency department visit rate for the composite outcome of the average monthly rate of initial emergency department visits for suicidal ideation, intentional self-poisoning, and intentional self-harm. Secondary outcomes were average monthly rates of death or intensive care unit admission resulting from index visit from April 2002 to December 2013. There was no significant change in emergency department visit rate for the composite outcome before and after announcement of Amanda Todd's death. Findings suggest publicity around this video release was not associated with increase in emergency department visits for suicidal behavior.
Singaravelu, 2015 [7], United Kingdom, high.	To identify and analyze websites potentially accessed by young people about self-harm	Searched only for websites where target audience were described as young people. Initial search terms were developed in discussion with 6 members of a Child and Adolescent Mental Health Services users group aged 15-19 years.	Sites accessed by self-harm or suicide search terms were mostly positive or preventative in tone, whereas sites accessed by the term "ways to kill yourself" tended to have a negative tone. A total of 314 websites were included in analysis, with evocative images of self-harm found in 21.0% (66/314) of sites.

[a]LGBT+: Lesbian, gay, bisexual, transgender, and others.

[b]Analysis of the same dataset of YouTube videos with one study examining the content of videos themselves and the other analyzing the comments.

Table 2. Summary of included studies examining the perspectives of individuals on the impact of sharing and viewing web-based images or videos of self-harm.

Category (Author, year, country, quality score)	User perspectives		
	Aims and objectives	Participants or sample	Results
Jacob, 2017 [9], United Kingdom, high.	To explore how young people understand and use web-based images of self-harm using semistructured interviews	21 individuals aged 16-24 years, living in Wales, United Kingdom, with a previous history of self-harm. Mean age for self-harm commencement was 13 years. A total of 16 participants sought professional help, 8 presented to emergency departments for their injuries; 86% (18/21) were female.	Some individuals cited the internet as a catalyst in the development of self-harm, where individuals were searching for advice and support and self-harm had "just come up" in the search with instructions and images. The majority engaged with web-based spaces to support and further develop a preexisting set of self-harming practices and reported the role of the internet in normalizing self-harm. Image rather than text-based interactions were the primary reason cited for using the internet for self-harm–related purposes. Images were said to invoke a physical reaction and inspire behavioral enactment. Participants reported viewing self-harm images as part of a ritualistic practice.
Seko, 2015 [38], multiple countries, high.	To conduct qualitative analysis of web-based interviews with individuals who produce self-harm content. To understand why content creators create and share self-harm–themed content and what needs are met by doing so	Creators of self-harm creative content, n=17; 82% (14/17) female	Thematic analysis of participants' narratives identified 2 prominent motives: self-oriented motivation (to express self and creativity, to reflect on self-harm experience, and to mitigate self-destructive urges) and social motivation (to support similar others, to seek out peers, and to raise social awareness). Participants also reported a double-edged impact of self-harm content both as a trigger and a deterrent to self-harm.
Sternudd, 2012 [39], multiple countries, moderate.	To examine discourses about self-injury photos from a user's perspective using web-based questionnaires	Users of self-injury forum: n=52; 87% (45/52) female	Informants reported that viewing or sharing images had alleviating rather than triggering effects with production of images often about memory and proof. Publishing them was seen as a way of sharing experiences with others and to give or receive help. Self-injury photos were described as a resource of a self-harm community culture. Informants often emphasized that the outcome of viewing these photos varies due to individual and situational differences.

Table 3. Summary of included studies examining web-based interventions using videos.

Category (Author, year, country, quality score)	Web-based video interventions		
	Aims and objectives	Participants or sample	Results
Choi, 2016 [40], United States, moderate.	To determine the feasibility of using a specific video in a web-based suicide awareness program for Asian American and non-Hispanic White college students	University students: n=431; 78.0% (336/431) female	Asian Americans rated the suicide awareness video significantly lower for cultural relevance than non-Hispanic Whites. Collectivist cultural orientation was a significant predictor for cultural relevance, credibility, and appeal. Cultural orientation and race or ethnicity should be strongly considered when web-based suicide awareness programs are developed for college students.
Park, 2014 [41], United States, high.	To determine the predictors of watching a web-based suicide prevention video and whether data characteristics differed for those watching most or only part of the video	University students: n=650; 71.8% (467/650) female	When examining characteristics of individuals who watched a suicide prevention video (which included self-harm content), the video completion group included more females, undergraduates, and Asian Americans and had higher individualistic orientation and more correct manipulation check answers. The video noncompletion group skipped items in a purposeful manner, showed less interest in the video, and spent less time completing questionnaires.
Robinson, 2015 [42], Australia, moderate.	To examine the safety and acceptability of a web-based suicide prevention program and determine which components were found to be most helpful and enjoyable	Secondary school students: n=21; 81% (17/21) female	A total of 21 young people completed the intervention. Overall, the intervention did not lead to increases in suicidal ideation or distress. Participants reported enjoying the program, in particular watching the video diaries and completing the activities. Most participants said they would recommend the program to a friend.
Scherr, 2017 [43], Germany, moderate.	To examine the impact of a suicide awareness video on adherence to newspaper reporting guidelines; video intended to be used for web-based format	Journalism students: n=78; 69% (54/78) female	Awareness material exposure helped to improve responsible reporting of suicide, with the awareness video showing a stronger effect than written material alone.

Categories of Studies

In total, 12 studies (13 articles) reported an analysis of images or videos related to self-harm or suicidal behaviors over a range of platforms (multiple sources, n=2; YouTube, n=6; Instagram, n=2; Tumblr, n=1; and Twitter, n=1). Three studies examined the perspectives of individuals on the impact of sharing and viewing web-based images or videos of self-harm [9,38,39]. Furthermore, 4 studies (1 high and 3 moderate quality) examined web-based interventions using videos [40-43]. All 4 studies reported the potential for positive impacts. Some studies [31] made a distinction between photographs of nonsuicidal self-harm and suicide attempts, but the criteria for these distinctions were often unclear, so the term self-harm is used throughout.

Use of Hashtags

Moreno et al [36] examined the nature of ambiguous self-harm–related hashtags (eg, #selfharmmm). At the time of this study, Instagram's terms of use (since revised [21]) blocked users from searching for the hashtag #selfharm. As a result, a number of ambiguous hashtags emerged to bypass the restrictions. The hashtag #selfharmmm was used to identify a number of other ambiguous hashtags linked to self-harm-related content. Such hashtags included #blithe, #selfinjury, and #ehtilb. The number of search results for each hashtag grew substantially over 150 days. Search results revealed a wider group of hashtags #mysecretfamily present across platforms, referring to a range of mental health issues, of which #cat represented self-harm.

Ethical Approvals

Ethical approval was stated as unnecessary in 5 of the 13 articles examining content across platforms because of the public nature of the internet sources. Of the 13 articles, 3 articles stated that they had received institutional review board ethical approval and only 6 included any discussion of ethical issues, such as protecting individual identities or sensitivity in disseminating results, although the content was likely to have been created by potentially vulnerable children and young people.

Platforms

Image-Based Platforms

Two studies examined self-harm content on Instagram [31,36]. Both found considerable image-related self-harm content. One examined photographs of self-harm over a 4-week period, investigating the nature of the pictures, associated comments,

and timing of posting behavior [31]. Most photographs depicted wounds with mild or moderate severity. Pictures were generally posted in the evenings. Most comments on self-harm–related posts were part of a more general discussion (3291/6651, 49.47%), with 11.58% (770/6651) asking the user to stop self-harming [31]. One study examined Tumblr posts [32]. Posts that were comforting, supportive, or related to prevention made up 8.00% (220/2750) of all posts. Posts interacting with other users made up 9.05% (249/2750), of which 47.0% (117/249) provided emotional support and 51.0% (127/249) sought or provided advice. Of those seeking or providing advice, 40.9% (52/127) provided positive or supportive advice, whereas 25.1% (32/127) provided potentially harmful advice, including advice on concealment.

One qualitative interview study found that Tumblr was often reported as the preferred platform because of the ease of searching for and sharing of images, perceived anonymity, and lack of moderation compared with other social media sites. This freedom to access and post severe images of self-harm was reported to lead to the normalization and exacerbation of self-harm for some participants [9].

Video-Based Platforms

Seven studies (8 articles) [2,6,28,30,32-35] analyzed YouTube videos, of which 3 (4 articles) focused on self-harm–related videos [2,6,28,35]. Although visual representations of self-harm were common, two-fifths of YouTube videos in one study exhibited a message encouraging help seeking [2]. Analysis of 100 videos found that 53.0% (53/100) were educational, often discouraging self-harm (26/100, 26.0%) [6]. Comments on this set of videos mostly shared personal experiences (38/100, 38.0%) [28]. There were no significant differences in the types of comments based on the number of views, ratings, or types of videos. An examination of videos concerning self-harm first aid tips found 28% (11/40) of videos included a disclaimer indicating that self-harm was *acceptable* providing certain safety precautions were followed [35]. Other YouTube videos varied in content, including *fire challenge* videos in which an individual doused themselves in a flammable liquid, set themselves alight, and attempted to extinguish the flames before serious burns are inflicted [29]. This was 1 of only 2 studies [29,37] included in this review in which the majority of participants were male (45/50, 90%).

An analysis of Twitter posts found images and videos related to self-harm, but these were not graphic in nature [34] and included images of celebrity tattoos covering self-harm scars with the message *stay strong*. These images were shared with positive messages of growth and recovery. Video links tended to be of individuals sharing their own stories.

Graphic Imagery

Images rather than text-based interactions have been reported as the primary reason for using the internet for self-harm purposes in semistructured interviews [9]. The nature of images was important in terms of outcomes. It was commonly expressed that *bluntly gruesome* photos [39] (eg, flesh, open wounds, and blood) were more triggering than pictures of scars and healing wounds [38]. Results from internet searches varied based on the search terms used, with the term *ways to kill yourself* eliciting a high proportion of sites with evocative images (20/43, 47%) [7]. Platforms vary according to the degree of graphic imagery. Visual representations of self-harm were common on YouTube. The majority of videos showed severe and open wounds and acute scarification, and 1 video portrayed 3 clips of suicide attempts [2]. Analysis of 100 videos similarly showed that images of self-harm were common (64/100, 64.0% of videos), with 28% (14/50) of character videos depicting live acts of self-harm [6]. A large number of self-harm images are posted daily on Instagram. Of 2826 self-harm images examined in one study, 39.60% (1119/2826) were rated as mild in severity (ie, superficial scratches), 47.78% (1350/2826) as moderate (ie, deeper cut and blood flowing), and 12.60% (356/2826) as severe (ie, very deep cut or a large amount of deeper cuts and blood), with 93.06% (2630/2826) of images depicting cuts [31]. In another study, 17.7% (127/718) of Tumblr posts on self-harm or suicide were graphic images or video clips [32].

A range of emotional reactions and impacts on the viewing or sharing of images or videos are described in Table 4.

Table 4. Summary of emotional reactions and impacts of viewing and sharing videos and images.

Impact of image or video	Findings reported
Anger or hostility	• Spoof advertisement on Twitter for stick-on self-harm scars evoked reactions of anger and frustration [34] • Some hostile comments about an uploader of self-harm content (57/864, 6.6%) were found in comments of YouTube videos [28] • A small percentage of comments of self-harm–related Instagram posts were coded as abuse (450/6651, 6.77%) [31]
Other emotions	• The reaction from viewing self-harm photographs varied considerably between informants and may be dependent on the individuals' state of mind when they are viewed. A wide range of feelings were covered in responses, including being sad, sick, and shocked. Reactions such as depression, grief, and concern for themselves were stated [39] • Common reactions reported after viewing a suicide prevention video included sadness, surprise, shock, and feeling overwhelmed. Almost 40% of participants indicated that they were most affected by the real, personal stories of family and friends of those who had taken their own lives (particularly the impact of suicide on the lives of the people left behind) or individuals who had survived a suicide attempt [40]
Ambivalence	• Dramatic responses are not always reported when viewing self-injury photos, with some reporting it can be done to pass time [39]
Exacerbation of self-harm urges or behavior	
Triggering	• Nearly 3 quarters of interview participants reported that imagery (notably photographs) was the primary reason for their utilization of the internet, due to a powerful physical reaction that triggered the desire to self-harm. Reliance on the image as a trigger had led to images assuming a vital role within their ritualistic practice with "sessions" often commencing with retrieval of a web-based image. The power of the image primarily centered on their ability to *"bring back memories"* of previous self-harm or the ease with which they allowed the individual to envisage how others experience the act. Participants also reported looking at images deliberately to trigger more severe self-harm [9] • Participants often discussed the triggering role of images. Many stated that whether an image would trigger an act depended on mood at the time. It was commonly expressed that photographs of flesh, open wounds with blood are more triggering than pictures of scars and healing wounds [38] • About one-third of participants describe the outcome of viewing photographs as triggering, with "bluntly gruesome" photographs described as the most triggering content [39]
Competition	• Participants spoke of being inspired to recreate certain sets of practices presented by particular images. Discussion was characterized by a sense of competition with individuals desiring to emulate the depicted self-harm while chiding themselves when they failed to engage with more sophisticated and severe techniques [9]
Imitation	• Pictures depicting wounds generated around twice as many comments from users than pictures not depicting wounds. There was also a significant association between wound grade and number of comments [31]. Time-related analyses did not support any effects of contagion or reinforcement [31,37] • Participants reported that a lack of moderation on Tumblr and the freedom to view and share the most stark and severe images had led to normalization and exacerbation of self-harm. One participant stated that their self-harm had escalated from little gashes to severe injuries and cutting through arteries [9]
Reduction in self-harm urges	
Calming	• Participants reported self-harm photographs as providing a sense of vicarious relief and of viewing photos to calm themselves when feeling triggered [38]. Reactions to self-harm images were described as comforting or calming in nearly half of statements [39]
Use as a deterrent	• Some participants stated that self-harm photographs of severe injuries acted as a deterrent of self-harm. Participants reported using this as a pre-emptive strategy to avoid more severe self-injury [38]. One participant described viewing of severe injuries as making them feel nauseated, serving as a strategy to avoid more severe self-harm [38]
Emotional outlet	• Content creation, particularly artistic or creative content, was described as an emotional outlet to disclose negative emotions, distress, and aspects of the self otherwise difficult to express. This was reported, at times, to reduce self-harm urges by acting as a distraction or alternative to self-harm [38] • Some participants reported that looking at content helped them reflect on their experience, make sense of it, and potentially avoid further episodes. Creation of content was reported to reduce self-harm urges serving as a creative alternative [38]. Participants reported viewing images made them feel less alone helping to curb NSSI[a] urges. Feelings of relief were reported with photographs reducing urges to self-harm.
Connection with others	

Impact of image or video	Findings reported
Empathy	• Empathetic comments made up 23.49% (1562/6651) of comments on Instagram posts [31]. Many participants spoke of feeling empathy with content creators when viewing images related to self-harm. Several participants also addressed the internet as their only source of support or connection where they could receive empathetic reactions and emotional support [38]. Empathy and sympathy were common reactions to a suicide prevention video [40]
Solidarity and reduction in loneliness	• Participants reported the viewing of photographs as comforting, as they made them feel less alone [38,39]. Participants describe feeling less alone in their battle as a motivation for sharing images [39]
Giving and receiving help	• The motivation to support like-minded people was often described as going hand in hand with a desire for help [38] • Warnings asking user to stop behavior were present in 11.58% (770/6651) comments on Instagram posts, as were offers of help (462/6651, 6.95%) [31] • A total of 51.0% (127/249) of interactive Tumblr posts involve seeking or providing advice, of which 40.9% (52/127) provided positive support or advice (eg, encouragement in stopping self-harm) and 25.1% (32/127) provided potentially harmful advice (eg, advising how to secretly engage in self-harm), with 13.4% (17/127) suggesting professional help or therapy [32] • An equal number of comments on YouTube videos asked for help or offered help (23/1150, 2.00%), and a small number encouraged the uploader to seek help (21/1150, 1.82%) [28] • Participants who watched a suicide prevention video expressed a higher awareness of the need to watch for signs of depression to be able to help friends and the need to take immediate action, take depression seriously, and talk openly about suicide [40] • A total of 9.09% (249/2739) of Tumblr posts involved directly interacting with another user. Of these, 47.0% (117/249) provided emotional support or reassuring messages to one another [32]
Feedback or discussion of creative content	• Creators of creative content expected constructive criticism from their viewers to improve their artistic skill, and positive feelings were reported when content received comments or was reblogged [38] • High levels of feedback were given or received for video content. Feedback included admiration of video quality (191/869, 22.0%) and the video message (148/869, 17.0%) and validation and admiration for the individual who uploaded the video (134/869, 15.4%) [28]. Very few Instagram comments complimented the wound or image (33/6651, 0.50%) [31]

[a]Nonsuicidal self-injury.

Negative Reactions and Impacts

Anger and hostility in response to images or videos were reported in 3 studies [28,31,34]. Other emotional reactions included sadness, shock, and concern [39,40]. Some have described ambivalence or simply viewing photographs to pass time [39]. Images having the potential to exacerbate self-harm through triggering, normalization, competition, or imitation have also been described [9,38,39].

Triggering

The triggering role of imagery was reported in 3 qualitative interview studies [9,38,39]. A third of the participants reported that photographs triggered self-harm urges [39]. Participants reported viewing images as part of a ritualistic practice to be triggered before harming themselves to self-harm more severely [9]. Both triggering and alleviating effects of images have been reported [39]. Whether the material was triggering was dependent on the viewer's mood. If an individual was already determined to self-harm, then looking at images would encourage urge, whereas if they were in a positive mood, the images would have minimal impact [38].

Normalization and Competition

There were reports of web-based engagement exacerbating self-harm due to normalization and exposure [9]. A sense of competition was reported in all 3 qualitative interview studies [9,38,39], with individuals desiring to emulate the depicted harm while chiding themselves or receiving negative comments from others when they failed to engage with more sophisticated and severe techniques [9]. Participants reported negative feelings regarding failure to harm themselves as severely as the injuries shown in photographs [38]. There were also reports of needing to make wounds worse to be a valid *self-injurer* [39]. A strong correlation between male informants and negative statements was found in 1 study, with 80% (4/5) of the statements expressing competitive reactions [39].

Imitation

Brown et al [31] examined several possible markers of imitative behavior on Instagram, including comments on photographs and a time-related analysis of posts. Pictures directly depicting wounds generated twice as many comments as those not depicting wounds [31]. A significant association was also observed between wound grade and number of comments, which could indicate a socially reinforcing function of posts. However, the time-related analysis of images did not indicate any effects of imitation or reinforcement [31]. A separate time series analysis did not find an increase in average monthly emergency department visits for suicide-related diagnoses in Ontario children aged 11-17 years following posting and publicity of a young person's YouTube video who took their own life [37]. This showed the young person holding up cards telling a story of bullying, mental health issues, and self-harm, with an image of self-harm injuries at the end.

Positive Reactions and Impacts

The production of images was often about memory and proof, with photos forming part of a narrative likened to fading scars. Individuals reported taking photographs for artistic value or their own interest [39]. A reduction in self-harm urges was reported with content creation said to act as an alternative outlet for negative emotions, with the viewing of images acting as a placebo or deterrent [38]. Those who create drawings, poems, and videos, in particular, called the materials a form of art that portrays the feelings behind self-harm, with positive feelings reported when content was reblogged [38].

Reactions of empathy, sympathy, and where people both gave and received help were found across various platforms [31,38,40]. Self-harm photographs were described as a resource for the self-harm community and were said to provide a feeling of solidarity and reassurance of not being alone [38,39].

Prevention and Intervention

Trigger Warnings

Three studies (4 articles) exploring the content of images or videos included an analysis of trigger warnings. These are placed on content by either the uploader, moderator, or platform and are intended as a warning to individuals who may find that the images increase the urge to self-harm. Two articles (based on a single data set of 100 videos) reported that 58.0% (58/100) of YouTube videos related to self-harm did not warn about content [6,28]. Just 2 of 40 self-harm first aid videos contained such a warning [35]. Two peer-moderated and 3 professionally driven sites were examined in an assessment of self-harm across different platforms. Both peer-moderated sites contained trigger warnings. The professionally driven websites did not contain any graphic material, so no trigger warnings were present. Trigger warnings were present on 4 of the 5 YouTube videos included in this study. Of the 5 Myspace groups examined, 3 contained trigger warnings and 2 prohibited self-harm images. Only 1 of the 4 Facebook groups contained a trigger warning, despite the presence of photos and videos in most groups [2].

Alongside user-generated trigger warnings are content advisory warnings on some platforms. Searches for self-harm-related content on Tumblr first provided a screen with suggestions for seeking help or finding more inspirational content, with the option to proceed to the next screen to view the search results [32]. Of the 18 self-harm–related hashtags, 6 generated a warning label on Instagram [36].

Recovery-Oriented Content

Recovery-oriented content was observed in varying degrees across platforms. Among Tumblr posts providing advice, only 13.4% (17/127) suggested professional help or therapy [32], and 2.4% (21/869) of comments on YouTube videos encouraged the uploader to seek help [28]. Some comments indicated that the individual was seeking treatment (83/869, 9.6%), had recovered (62/869, 7.1%), or expressed a desire to recover (33/869, 3.7%) [28]. Recovery-oriented content appeared to be most evident on Twitter, with inspirational quotes and stories of recovery [34].

Interventions

The interventions discussed here include a web-based video component to an intervention designed to reduce suicide or self-harm. The content of these videos was professionally generated and differed from the user-generated content discussed above and contained no graphic self-harm imagery.

The safety and acceptability of a web-based suicide prevention program incorporating video elements in secondary school students was examined in one study [42]. The video component of this intervention uses a *host* character to deliver verbal therapy and parts of this include *video diaries* that tell a story portrayed with young people actors. Overall, the intervention did not lead to increases in suicidal ideation or distress, and 71.4% (15/21) of participants rated the video diary component as more enjoyable and helpful than hearing from website moderators or receiving text messages [42]. The impact of a suicide awareness video as part of a web-based intervention to increase adherence to newspaper reporting guidelines in journalism students was examined in another study. Exposure to awareness raising material with or without the video helped to improve responsible reporting of suicide, with the awareness raising video showing the strongest effect [43].

Two studies examined different aspects of *The truth about suicide* suicide awareness video. This video focused on interviews with people who have lost a loved one to suicide and the testimonials of students who have experienced depression and suicidal ideation. Both studies found video streaming to be a feasible method for delivering a suicide awareness program to college students [40,41]. Predictors of watching the 29-min video to completion were female gender, undergraduate, Asian ethnicity, and higher individualism as opposed to collectivism [41], measured using the individualism-collectivism questionnaire [44]. Asian American students rated the video significantly lower on cultural relevance than non-White Hispanic students [40]. Following viewing of the video, participants expressed a higher awareness of the signs of depression, the need to take immediate action, and to talk openly about suicide [40]. Small increases in distress were found [41]. The importance of an effective debriefing session in a safe and confidential setting for sharing thoughts, feelings, and experiences and to prevent iatrogenic effects of web-based suicide prevention programs using emotionally charged videos was emphasized and successfully delivered on the web [40].

Discussion

Principal Findings

We conducted a systematic review of the literature related to the impact of sharing or viewing web-based self-harm images or videos, exploring intent and interpretation in young people. A total of 19 studies (20 articles) were included. No previous studies have focused on this area [18,19]. We have identified specific features of platforms that contribute to potentially harmful use and of images that may influence the impact on the viewer or sharer.

In recent years, there has been a move away from the once popular self-harm forums to other platforms in terms of use by

both individuals and researchers [9,45]. There has also been an increase in graphic imagery identified by self-harm–related searches over time [11], and participants reported image-based rather than textual interactions as the primary reason for internet use [9]. Studies consistently report a large presence of graphic imagery of self-harm on image-based platforms such as Instagram and Tumblr. In contrast, there is less graphic imagery present on Twitter and a greater presence of recovery-oriented content [34]. The difference between platforms was further supported by an analysis of randomly generated posts found with the hashtag #cutting. Instagram posts were found to display the greatest proportion of graphic content, with the smallest proportions found on Twitter [46]. Recovery-oriented resources were congregated on the platforms with the least problematic content. Potentially harmful content appeared to congregate on the platforms that have relatively little moderation, where participants can remain anonymous and search for images easily [9]. The use of videos in suicide prevention interventions has also been examined [40-43]. The content of these videos differed from user-generated materials, such as those shared on social media platforms. Videos included as part of an intervention focused on issues around suicide or self-harm, including interviews with those impacted by suicide [40,41], video diaries delivered by actors [42], and suicide awareness material related to response reporting of suicide by the press [43].

A range of reactions and intentions have been reported in relation to posting or viewing images of self-harm from empathy, solidarity, and the use of images to give or receive help [31,38] to potentially harmful advice suggesting new methods and tips for hiding self-harm, as well as normalization and exacerbation of self-harm [9,39]. This is further supported by recent research describing strong social motivations for posting self-harm content on Instagram as well as the need to document self-harm or recovery [47]. Viewing images as an alternative to self-harm and as a creative outlet were regarded in 2 studies as positive impacts [38,39]. The use of images as a creative outlet for distress was found in a recent analysis across platforms [48]. Reactions of anger, hostility, and ambivalence in posting self-harm images have also been reported [28,31,34].

Self-harm images were reported to be part of a ritualistic practice by all 3 studies that examined user perspectives. Participants reported viewing graphic imagery on the web to evoke the right mood for self-harm, often resulting in more severe injuries [9]. Photographs of fading scars or creative content have been reported to have a positive impact [38,39]. Photographs of fresh cuts or severe injuries are most frequently reported to have a negative impact [9]. Participants reported that viewing more severe injuries leads to negative feelings of failure to achieve the same level of injury or more severe injuries over time [9,39].

There was some evidence of the role of imitation and reinforcement, driven partly by the number of comments generated by self-harm imagery and wound severity [9,31]. Although time series analyses were conducted in 2 studies [31,37], the limitations of these analyses indicate the need for further studies to examine the impact of self-harm imagery. Although the evidence for social contagion is lacking in the papers described here, recent studies relating to the release of the TV series *13 Reasons Why,* in which the suicide of a young female is graphically depicted, found this to be associated with a significant increase in adolescent suicide rates [49] concordant with the period in which social media discussions of the series were greatest [50]. This increase in deaths was found despite some individuals reporting positive impacts [51]. The graphic scene was removed following concerns expressed by the suicide prevention agencies. Recent research has suggested a need to think beyond a model of contagion of self-harm behavior and shifting the focus onto other mechanisms of harm and benefit [52]. It has been highlighted that social media often acts to mirror society and that a wider context needs to be taken. For example, gaps in service provision often led people to seek online peer support [52].

Strengths and Limitations

This study provides a comprehensive overview of studies directly examining web-based self-harm imagery or videos in relation to children and young people. The potential for publication bias exists in any literature review. Steps were taken to minimize bias, including conducting an extensive literature search of multiple databases, including topic-specific websites and reviewing reference lists, titles and abstracts being screened by 2 researchers, and data extraction being conducted by 2 researchers for each article. Only English language publications were included in the analysis. Females outnumber males, where this information was available, in all but 2 studies, one of videos related to the fire challenge [29] and a time series analysis of emergency department attendance [37]. One study reported that males were more likely than females to report negative impacts of self-harm images [39]. However, the limited number of males included makes it difficult to draw firm conclusions. Further research into potential sex differences and involving larger samples of males is needed.

Although a number of measures were included in eligible studies, including analyses of content, the impact of self-harm imagery has relied heavily on self-reporting of short-term impacts. This is particularly true for the perceived positive impacts. Future research should aim to understand the long-term effects of continual exposure to such images, as the nature of immediate impacts may change over time. This is highlighted by the recent literature related to *13 Reasons Why,* where an increase in suicide rates was found [50], despite some individuals reporting a positive impact of the show [51]. Web-based content has the potential to reach a large audience; as such, a range of emotional reactions is to be expected. It is important that the overall impacts on outcomes such as suicide rates and levels of self-harm are assessed alongside self-report measures to put any reported positive or negative impacts into a wider context. The use of validated outcome measures related to self-harm or levels of distress would strengthen the evidence base in this area.

Although 2 studies included a time series analyses [31,37], both had their limitations. It is not yet clear how and whether any long-term process of imitation would take place on a platform such as Instagram, where images can be searched and viewed long after they are posted [31]. Although the potential impact of a particular video was examined for emergency department attendances, this was confined to 1 region that was not where

the death occurred and media dissemination was greatest. This would not have been a measure of self-harm that presented to settings other than emergency departments, such as primary care, or for self-harm in the community that would not be present in health care data. Further research is needed before any conclusions can be drawn based on such evidence. This could include the use of survey data or routinely collected health care data spanning multiple settings (eg, primary and secondary care). Accurately determining the characteristics of young people based on web-based profiles is not always possible. The average age of participants is likely to be younger than that reported, as young people often misrepresent their age to gain access to restricted content [6]. Finally, although self-harm presentations by young people to general hospitals most commonly involve self-poisoning [53], this method of self-harm rarely featured in the images appearing on the internet. Thus, the images that are found largely reflect self-harming behavior most frequently occurring in young people in the community [53], although accessing internet sources before self-harm is common in young people who attend hospital following self-poisoning [54].

Implications

Clinicians working with young people who self-harm should routinely enquire about internet use [54], support them in recognizing and managing triggering content, and encourage healthier web-based behaviors. The powerful role of imagery in evoking emotional reactions and motivating both potentially harmful behaviors, such as self-harm [16], and adaptive behavior [55] could make this an important factor in recovery.

As young people frequently create their own content, guidance to increase safe and responsible depiction of suicide and self-harm image and video-based content are now needed to educate young people as well as media professionals. Recent analysis of videos related to the *Blue Whale Challenge* on YouTube has demonstrated that although these user-generated videos seek to raise awareness, most videos violate guidance around safe and effective messaging, further underscoring the need to educate social media users and content generators on safe content creation and the factors that may have either positive or negative impacts on vulnerable viewers [56]. Individuals have reported an awareness of the potential impacts of self-harm images and videos and as a consequence add trigger warnings to their own posts [9]. Initiatives could be developed to build on this and to equip young people with the skills to analyze and evaluate their own creative content and to highlight the potential for posts intended as catharsis and therapy for themselves to be potentially harmful to vulnerable individuals. Strategies to improve safety could build on the work of the #chatsafe project and the development of guidelines for safe peer-to-peer web-based communication related to suicide [57]. Such projects could be used to inform parents, teachers, and other caregivers as well as young people. This could improve the safety of image-based platforms.

A small number of studies examined the use of videos disseminated on the web as a means of intervention. Given the large self-harm community using image and video-based platforms, further research into potential interventions is needed to understand how to maximize effectiveness. Content creation can act as an alternative outlet for negative emotions, and the viewing of images acting as a placebo or deterrent could be explored as potential interventions. It has been suggested that educational programs should be developed to inform young people about the impact of the content of posts and how to respond to posts from those in distress [2,57]. A recent pilot study found that hopeful reactions to the self-harm content of YouTube can improve positive attitudes toward recovery [58]. This has the potential to prevent suicide contagion via social media. The use of videos in suicide prevention programs appeared to be promising not just for young people themselves but also for improving responsible reporting of suicide among journalism students [43]. The importance of the cultural relevance of prevention videos has been emphasized [41]. More recent coproduction of suicide prevention material with a popular YouTuber in Hong Kong suggests that this may be an effective avenue for future suicide prevention strategies [59].

Individuals reported an awareness of the potential impacts of self-harm images and videos and added trigger warnings to their own posts [9]. Automated platform content warnings may not have been sufficient [36]. Our results should be discussed in the context of recent changes in policy by platforms, including Facebook [60] and Instagram [21], regarding the posting of graphic self-harm imagery. Changes were made following the death of a young person in the United Kingdom in February 2019 and the subsequent campaign by the family. Instagram announced a ban on all images of self-harm, including pictures showing scars [21]. The particular features of platforms that make them more amenable to harmful use include a lack of moderation and ease of sharing, searching, and viewing of images [9]. Monitoring and regulation of posts may be a positive move toward making these spaces safer. The effectiveness of such policies has yet to be evaluated, and a recent analysis of YouTube content found that content did not comply with YouTube's policy on self-harm. Most content did not require age registration, and the specific help-related search terms had to be used to access any suicide prevention–related content [61]. Young people often find ways around posting restrictions, such as by using ambiguous hashtags [36] or moving to more hidden parts of the internet. An examination of 1 of these ambiguous hashtags on Instagram has shown that 3 quarters of posts contained a wound, the majority of which did not contain an advisory warning [62]. The use of image recognition software to identify graphic self-harm imagery could have important implications for protecting vulnerable individuals from potentially triggering content [63].

The potential for posting restrictions to reduce positive impacts, such as support and a sense of community, must also be considered as should any potential negative impacts of the removal of content posted by vulnerable young people without consultation. The use of social media to bridge support in service provision, particularly with regard to both giving and receiving peer support, has been further highlighted in recent literature, pointing to the potential inadequacy of a strategy focused on placing pressure on internet service providers to remove all self-harm content. Instead, care must be taken with such restrictions to avoid further harm [52]. Banning images of scars

may have the unintended consequence of increasing stigma or removing from individuals the opportunity to share their stories, including those related to recovery. The indiscriminate removal or blurring of images of people with injuries and scars will include those not related to self-harm and could negate the potentially positive impact of fading or healing self-harm scars. Although the results of this review support concern related to safety and exacerbation of self-harm, the potential for positive impacts should not be underestimated. Future research should seek to evaluate the effectiveness of current posting restrictions and to identify the best strategies to reduce risk and maximize positive impacts on participants by incorporating user perspectives. This supports previous calls to better understand how to leverage the unique opportunities afforded by such platforms to reach and engage vulnerable individuals [46,47], with the potential for awareness materials to reach a large audience of vulnerable individuals via social media [64].

We identified no studies examining the role of the Google Images search engine, which has not seen the same restrictions applied to searches for self-harm imagery as social media platforms. There is no social community or moderation of Google images. Google Images does not incorporate the same tools that exist on the main Google search engine, ensuring results related to helplines, and information appears at the top of search results for suicide and self-harm. Typing *self-harm* into Google images results in a collection of images from across other sites. This is a powerful tool and appears not to have been the focus of this kind of research or policy intervention. It remains an ongoing challenge for the literature to remain current with the changing popularity of various platforms, for example, with the current popularity of Tik Tok, which does not appear to be the focus of current research.

It is important to note that the majority of studies reported here explore immediate or short-term reactions to self-harm content. Future research should aim to understand the long-term effects of continual exposure to such images, as the positive or negative immediate impacts may change over time.

Careful consideration should also be paid to the ethical issues of conducting research on freely available content and the associated implied consent. Future research should ensure the recruitment of males and focus on delineating the specific features of images that may be harmful and the mechanisms involved.

Conclusions

The way in which young people use the internet is continually evolving. An increasing preference is being reported for the use of the internet for image-based rather than text-based interaction with regard to self-harm. Concerns over negative impacts such as exacerbation and normalization of self-harm and sharing of information on new methods or concealment of self-harm are supported by research. However, there can also be positive impacts such as seeking and receiving peer support and viewing images as an alternative to self-harm and as an outlet for negative feelings. Graphic images and photographs of severe injuries are most often reported to have a negative impact. Clinicians assessing distressed young people should routinely enquire about internet use, explore strategies to manage any triggering content, and be aware of helpful sites. The use of images and videos as part of web-based interventions for suicide prevention is an area worthy of further study. Future research should also seek to evaluate the effectiveness of current posting restrictions on social media and to educate clinicians and caregivers on how to encourage healthier web-based behaviors. The combination of appropriate policies by internet service providers and education for individuals may be necessary to allow for the potential positive impacts of web-based imagery while minimizing the potential for harm.

Acknowledgments

KH is a National Institute for Health Research Senior Investigator and is supported by the Oxford Health NHS Foundation Trust. This work was funded by the Health and Social Care Research Wales (grant number SCS-14-11) and the MQ Mental Health Research Charity through the Adolescent Mental Health Data Platform (ADP; Grant Reference MQBF/3 ADP). The views expressed are entirely those of the authors and should not be assumed to be the same as those of ADP or MQ Mental Health Research Charity. The funders had no role in the study design; collection, analysis, or interpretation of data; writing of the report; or the decision to submit the paper for publication.

Authors' Contributions

AJ and AM conceived the study. All the authors contributed to the study design. Literature searches, data extraction, and analysis were conducted by AM, LB, and AJ. AM and AJ drafted the manuscript. All authors commented on the manuscript.

Conflicts of Interest

AM reports grants from Health and Social Care Research Wales. Funders did not have any role in study design, conceptualization, interpretation of results, or decision to publish. AJ reports grants from Health and Social Care Research Wales and grants from MQ during the conduct of the study. Funders did not have any role in study design, conceptualization, interpretation of results, or decision to publish. KH, AS, and LB have no conflicts of interest to disclose.

Multimedia Appendix 1
Data extraction form.

[DOCX File , 16 KB - jmir_v23i3e18048_app1.docx]

References

1. Mitchell KJ, Ybarra ML. Online behavior of youth who engage in self-harm provides clues for preventive intervention. Prev Med 2007 Nov;45(5):392-396. [doi: 10.1016/j.ypmed.2007.05.008] [Medline: 17599399]
2. Lewis SP, Heath NL, Michal NJ, Duggan JM. Non-suicidal self-injury, youth, and the Internet: what mental health professionals need to know. Child Adolesc Psychiatry Ment Health 2012 Mar 30;6(1):13 [FREE Full text] [doi: 10.1186/1753-2000-6-13] [Medline: 22463379]
3. Castro TS, Osório AJ. 'I love my bones!'1 – self-harm and dangerous eating youth behaviours in Portuguese written blogs. Young Consumers 2013 Nov 15;14(4):321-330. [doi: 10.1108/yc-03-2013-00351]
4. Castro T, Osório A. Online violence: not beautiful enough… not thin enough; anorectic testimonials in the web. PsychNology Journal 2012;10(3):169-186 [FREE Full text]
5. Zdanow C, Wright B. The Representation of Self Injury and Suicide on Emo Social Networking Groups. African Sociological Review 2012;16(2):1027-4332.
6. Lewis SP, Heath NL, St DJ, Noble R. The scope of nonsuicidal self-injury on YouTube. Pediatrics 2011 Mar;127(3):e552-e557. [doi: 10.1542/peds.2010-2317] [Medline: 21339269]
7. Singaravelu V, Stewart A, Adams J, Simkin S, Hawton K. Information-Seeking on the internet. Crisis 2015;36(3):211-219. [doi: 10.1027/0227-5910/a000307] [Medline: 26088826]
8. Baker TG, Lewis SP. Responses to online photographs of non-suicidal self-injury: a thematic analysis. Arch Suicide Res 2013;17(3):223-235. [doi: 10.1080/13811118.2013.805642] [Medline: 23889572]
9. Jacob N, Evans R, Scourfield J. The influence of online images on self-harm: a qualitative study of young people aged 16-24. J Adolesc 2017 Oct;60:140-147 [FREE Full text] [doi: 10.1016/j.adolescence.2017.08.001] [Medline: 28881214]
10. Feierabend S, Plankenhorn T, Rathgeb T. (Multi-) Media Basisstudie zum Medienumgang 12-bis 19-Jähriger in Deutschland. JIM 2015 Jugend, Information, Stuttgart. 2016. URL: https://www.textilwirtschaft.de/news/media/8/JIM-Studie-2015-70869.pdf [accessed 2020-10-24]
11. Biddle L, Derges J, Mars B, Heron J, Donovan JL, Potokar J, et al. Suicide and the internet: changes in the accessibility of suicide-related information between 2007 and 2014. J Affect Disord 2016 Jan 15;190:370-375. [doi: 10.1016/j.jad.2015.10.028] [Medline: 26546772]
12. Libby LK, Shaeffer EM, Eibach RP, Slemmer JA. Picture yourself at the polls: visual perspective in mental imagery affects self-perception and behavior. Psychol Sci 2007 Mar;18(3):199-203. [doi: 10.1111/j.1467-9280.2007.01872.x] [Medline: 17444910]
13. Holmes EA, Mathews A. Mental imagery in emotion and emotional disorders. Clin Psychol Rev 2010 Apr;30(3):349-362. [doi: 10.1016/j.cpr.2010.01.001] [Medline: 20116915]
14. Crane C, Shah D, Barnhofer T, Holmes EA. Suicidal imagery in a previously depressed community sample. Clin Psychol Psychother 2012;19(1):57-69 [FREE Full text] [doi: 10.1002/cpp.741] [Medline: 21254309]
15. Ng RM, Di Simplicio M, McManus F, Kennerley H, Holmes EA. 'Flash-forwards' and suicidal ideation: a prospective investigation of mental imagery, entrapment and defeat in a cohort from the Hong Kong Mental Morbidity Survey. Psychiatry Res 2016 Dec 30;246:453-460 [FREE Full text] [doi: 10.1016/j.psychres.2016.10.018] [Medline: 27792974]
16. Hasking PA, Di Simplicio M, McEvoy PM, Rees CS. Emotional cascade theory and non-suicidal self-injury: the importance of imagery and positive affect. Cogn Emot 2018 Aug;32(5):941-952 [FREE Full text] [doi: 10.1080/02699931.2017.1368456] [Medline: 28838289]
17. McEvoy PM, Hayes S, Hasking PA, Rees CS. Thoughts, images, and appraisals associated with acting and not acting on the urge to self-injure. J Behav Ther Exp Psychiatry 2017 Dec;57:163-171. [doi: 10.1016/j.jbtep.2017.05.010] [Medline: 28601695]
18. Daine K, Hawton K, Singaravelu V, Stewart A, Simkin S, Montgomery P. The power of the web: a systematic review of studies of the influence of the internet on self-harm and suicide in young people. PLoS One 2013;8(10):e77555 [FREE Full text] [doi: 10.1371/journal.pone.0077555] [Medline: 24204868]
19. Marchant A, Hawton K, Stewart A, Montgomery P, Singaravelu V, Lloyd K, et al. A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown. PLoS One 2017;12(8):e0181722. [doi: 10.1371/journal.pone.0181722] [Medline: 28813437]
20. Rodham K, Gavin J, Lewis S, Bandalli P, St. Denis J. The NSSI Paradox: Discussing and Displaying NSSI in an Online Environment. Deviant Behavior 2016 Apr 21;37(10):1110-1117. [doi: 10.1080/01639625.2016.1169747]
21. Taking More Steps To Keep The People Who Use Instagram Safe. Instagram. URL: https://about.instagram.com/blog/announcements/more-steps-to-keep-instagram-users-safe [accessed 2019-11-20]
22. NICE. Self Harm: Longer-Term Management. NICE Clinical Guideline, No. 133. National Collaborating Centre for Mental health 2012:1-30. [Medline: 31886959]
23. Popay J, Roberts H, Snowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme 2006;1:1-87. [Medline: 27466647]
24. CASP. Critical Appraisal Skills Programme. 2013. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2019-04-11]

25. Darlow B, Fullen BM, Dean S, Hurley DA, Baxter GD, Dowell A. The association between health care professional attitudes and beliefs and the attitudes and beliefs, clinical management, and outcomes of patients with low back pain: a systematic review. Eur J Pain 2012 Jan;16(1):3-17 [FREE Full text] [doi: 10.1016/j.ejpain.2011.06.006] [Medline: 21719329]
26. Feder GS, Hutson M, Ramsay J, Taket AR. Women exposed to intimate partner violence: expectations and experiences when they encounter health care professionals: a meta-analysis of qualitative studies. Arch Intern Med 2006 Jan 9;166(1):22-37. [doi: 10.1001/archinte.166.1.22] [Medline: 16401807]
27. Marchant A, John A. The relationship between online imagery and self-harm: a systematic review of the literature. CRD42019148695. 2019. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019148695 [accessed 2019-09-04]
28. Lewis SP, Heath NL, Sornberger MJ, Arbuthnott AE. Helpful or harmful? An examination of viewers' responses to nonsuicidal self-injury videos on YouTube. J Adolesc Health 2012 Oct;51(4):380-385. [doi: 10.1016/j.jadohealth.2012.01.013] [Medline: 22999839]
29. Avery AH, Rae L, Summitt JB, Kahn SA. The Fire Challenge: A Case Report and Analysis of Self-Inflicted Flame Injury Posted on Social Media. J Burn Care Res 2016;37(2):e161-e165. [doi: 10.1097/BCR.0000000000000324] [Medline: 26862698]
30. Basch CH, Ruggles KV, Berdnik A, Basch CE. Characteristics of the most viewed YouTube videos related to bullying. Int J Adolesc Med Health 2015 Nov 10;29(4). [doi: 10.1515/ijamh-2015-0063] [Medline: 26556842]
31. Brown RC, Fischer T, Goldwich AD, Keller F, Young R, Plener PL. #cutting: Non-suicidal self-injury (NSSI) on Instagram. Psychol Med 2018 Jan;48(2):337-346. [doi: 10.1017/S0033291717001751] [Medline: 28705261]
32. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Connolly S, Rosas C, Bharadwaj M, et al. An analysis of depression, self-harm, and suicidal ideation content on tumblr. Crisis 2016 Jul 22:1-9. [doi: 10.1027/0227-5910/a000409] [Medline: 27445014]
33. Grzanka PR, Mann ES. Queer youth suicide and the psychopolitics of 'It Gets Better'. Sexualities 2014 May 13;17(4):369-393. [doi: 10.1177/1363460713516785]
34. Hilton C. Unveiling self-harm behaviour: what can social media site twitter tell us about self-harm? A qualitative exploration. J Clin Nurs 2016 Sep 07. [doi: 10.1111/jocn.13575] [Medline: 27604049]
35. Lewis SP, Knoll AKI. Do it yourself: examination of self-injury first aid tips on youtube. Cyberpsychol Behav Soc Netw 2015 May;18(5):301-304. [doi: 10.1089/cyber.2014.0407] [Medline: 25965864]
36. Moreno MA, Ton A, Selkie E, Evans Y. Secret society 123: understanding the language of self-harm on instagram. J Adolesc Health 2016 Jan;58(1):78-84 [FREE Full text] [doi: 10.1016/j.jadohealth.2015.09.015] [Medline: 26707231]
37. Poonai N, Mehrotra S, Mamdani M, Patmanidis A, Miller M, Sukhera J, et al. The association of exposure to suicide-related internet content and emergency department visits in children: A population-based time series analysis. Can J Public Health 2018 Jan 22;108(5-6):e462-e467. [doi: 10.17269/cjph.108.6079] [Medline: 29356650]
38. Seko Y, Kidd SA, Wiljer D, McKenzie KJ. On the creative edge: exploring motivations for creating non-suicidal self-injury content online. Qual Health Res 2015 Oct;25(10):1334-1346. [doi: 10.1177/1049732315570134] [Medline: 25662942]
39. Sternudd HT. Photographs of self-injury: production and reception in a group of self-injurers. J Youth Studies 2012 Jun;15(4):421-436. [doi: 10.1080/13676261.2012.663894]
40. Choi H, Park H, Suarez ML, Park C, Zhao Z, Wilkie DJ. Feasibility of a web-based suicide awareness programme for Asian American college students. BMJ Open 2016 Dec 21;6(12):e013466 [FREE Full text] [doi: 10.1136/bmjopen-2016-013466] [Medline: 28003296]
41. Park H, Choi H, Suarez ML, Zhao Z, Park C, Wilkie DJ. Predictors of valid engagement with a video streaming web study among asian american and non-Hispanic white college students. Comput Inform Nurs 2014 Apr;32(4):156-165 [FREE Full text] [doi: 10.1097/CIN.0000000000000035] [Medline: 24535027]
42. Robinson J, Hetrick S, Cox G, Bendall S, Yung A, Pirkis J. The safety and acceptability of delivering an online intervention to secondary students at risk of suicide: findings from a pilot study. Early Interv Psychiatry 2015 Dec;9(6):498-506. [doi: 10.1111/eip.12136] [Medline: 24684927]
43. Scherr S, Arendt F, Schäfer M. Supporting Reporting: On the Positive Effects of Text- and Video-Based Awareness Material on Responsible Journalistic Suicide News Writing. Arch Suicide Res 2017;21(4):646-658. [doi: 10.1080/13811118.2016.1222975] [Medline: 27602541]
44. Singelis TM, Triandis HC, Bhawuk DP, Gelfand MJ. Horizontal and vertical dimensions of individualism and collectivism: a theoretical and measurement refinement. Cross-Cultural Research 2016 Jul 25;29(3):240-275. [doi: 10.1177/106939719502900302]
45. Harris IM, Roberts LM. Exploring the use and effects of deliberate self-harm websites: an internet-based study. J Med Internet Res 2013;15(12):e285 [FREE Full text] [doi: 10.2196/jmir.2802] [Medline: 24362563]
46. Miguel EM, Chou T, Golik A, Cornacchio D, Sanchez AL, DeSerisy M, et al. Examining the scope and patterns of deliberate self-injurious cutting content in popular social media. Depress Anxiety 2017 Sep;34(9):786-793. [doi: 10.1002/da.22668] [Medline: 28661053]
47. Brown RC, Fischer T, Goldwich DA, Plener PL. 'I just finally wanted to belong somewhere'-Qualitative Analysis of Experiences With Posting Pictures of Self-Injury on Instagram. Front Psychiatry 2020;11:274 [FREE Full text] [doi: 10.3389/fpsyt.2020.00274] [Medline: 32372983]

48. Obermair HM, Dodd RH, Bonner C, Jansen J, McCaffery K. 'It has saved thousands of lives, so why change it?' Content analysis of objections to cervical screening programme changes in Australia. BMJ Open 2018 Dec 13;8(2):e019171 [FREE Full text] [doi: 10.1136/bmjopen-2017-019171] [Medline: 29440214]
49. Bridge JA, Greenhouse JB, Ruch D, Stevens J, Ackerman J, Sheftall AH, et al. Association Between the Release of Netflix's 13 Reasons Why and Suicide Rates in the United States: An Interrupted Time Series Analysis. J Am Acad Child Adolesc Psychiatry 2019 Apr 28. [doi: 10.1016/j.jaac.2019.04.020] [Medline: 31042568]
50. Niederkrotenthaler T, Stack S, Till B, Sinyor M, Pirkis J, Garcia D, et al. Association of increased youth suicides in the United States with the release of 13 reasons why. JAMA Psychiatry 2019 May 29. [doi: 10.1001/jamapsychiatry.2019.0922] [Medline: 31141094]
51. Nesi J, Johnson SE, Altemus M, Thibeau HM, Hunt J, Wolff JC. Perceptions and correlates of media influence in psychiatrically hospitalized adolescents. Arch Suicide Res 2020 Jun 26:1-12. [doi: 10.1080/13811118.2020.1779155] [Medline: 32589861]
52. Lavis A, Winter R. #Online harms or benefits? An ethnographic analysis of the positives and negatives of peer-support around self-harm on social media. J Child Psychol Psychiatry 2020 Aug;61(8):842-854. [doi: 10.1111/jcpp.13245] [Medline: 32459004]
53. Geulayov G, Kapur N, Turnbull P, Clements C, Waters K, Ness J, et al. Epidemiology and trends in non-fatal self-harm in three centres in England, 2000-2012: findings from the Multicentre Study of Self-harm in England. BMJ Open 2016 Apr 29;6(4):e010538 [FREE Full text] [doi: 10.1136/bmjopen-2015-010538] [Medline: 27130163]
54. Padmanathan P, Carroll R, Biddle L, Derges J, Potokar J, Gunnell D. Suicide and self-harm related internet use in patients presenting to hospital with self-harm: a cross-sectional study. Lancet 2016 Nov;388:S2. [doi: 10.1016/s0140-6736(16)32238-3]
55. Holmes EA, Arntz A, Smucker MR. Imagery rescripting in cognitive behaviour therapy: images, treatment techniques and outcomes. J Behav Ther Exp Psychiatry 2007 Dec;38(4):297-305. [doi: 10.1016/j.jbtep.2007.10.007] [Medline: 18035331]
56. Khasawneh A, Chalil Madathil K, Dixon E, Wiśniewski P, Zinzow H, Roth R. Examining the self-harm and suicide contagion effects of the blue whale challenge on youtube and twitter: qualitative study. JMIR Ment Health 2020 Jun 5;7(6):e15973 [FREE Full text] [doi: 10.2196/15973] [Medline: 32515741]
57. Robinson J, Bailey E, Hetrick S, Paix S, O'Donnell M, Cox G, et al. Developing social media-based suicide prevention messages in partnership with young people: exploratory study. JMIR Ment Health 2017 Oct 4;4(4):e40 [FREE Full text] [doi: 10.2196/mental.7847] [Medline: 28978499]
58. Lewis SP, Seko Y, Joshi P. The impact of YouTube peer feedback on attitudes toward recovery from non-suicidal self-injury: an experimental pilot study. Digit Health 2018;4:2055207618780499 [FREE Full text] [doi: 10.1177/2055207618780499] [Medline: 31463075]
59. Cheng Q, Shum AK, Ip FW, Wong HK, Yip WK, Kam AH, et al. Co-creation and impacts of a suicide prevention video. Crisis 2020 Jan;41(1):7-14. [doi: 10.1027/0227-5910/a000593] [Medline: 31066312]
60. Tightening Our Policies and Expanding Resources to Prevent Suicide and Self-Harm. Facebook. 2019. URL: https://about.fb.com/news/2019/09/tightening-our-policies-and-expanding-resources-to-prevent-suicide-and-self-harm/ [accessed 2020-10-23]
61. Niederkrotenthaler T, Schacherl R, Till B. Communication about suicide in YouTube videos: content analysis of German-language videos retrieved with method-and help-related search terms. Psychiatry Res 2020 Aug;290:113170. [doi: 10.1016/j.psychres.2020.113170] [Medline: 32526517]
62. Fulcher JA, Dunbar S, Orlando E, Woodruff SJ, Santarossa S. #selfharn on Instagram: understanding online communities surrounding non-suicidal self-injury through conversations and common properties among authors. Digit Health 2020;6:2055207620922389 [FREE Full text] [doi: 10.1177/2055207620922389] [Medline: 32426154]
63. Scherr S, Arendt F, Frissen T, Oramas M J. Detecting intentional self-harm on instagram: development, testing, and validation of an automatic image-recognition algorithm to discover cutting-related posts. Soc Sci Comput Rev 2019 Mar 18;38(6):673-685. [doi: 10.1177/0894439319836389]
64. Arendt F. Suicide on instagram - content analysis of a german suicide-related hashtag. Crisis 2019 Jan;40(1):36-41. [doi: 10.1027/0227-5910/a000529] [Medline: 29932019]

Abbreviations

ADP: Adolescent Mental Health Data Platform
CASP: critical appraisal skills programme
NHS: National Health Service

Edited by G Eysenbach; submitted 30.01.20; peer-reviewed by M Sinyor, J Robinson; comments to author 24.06.20; revised version received 24.10.20; accepted 17.01.21; published 19.03.21.

Please cite as:
Marchant A, Hawton K, Burns L, Stewart A, John A
Impact of Web-Based Sharing and Viewing of Self-Harm–Related Videos and Photographs on Young People: Systematic Review
J Med Internet Res 2021;23(3):e18048
URL: https://www.jmir.org/2021/3/e18048
doi:10.2196/18048
PMID:33739289

Review

Diabetology 4.0: Scoping Review of Novel Insights and Possibilities Offered by Digitalization

Claudia Eberle[1], Prof Dr, MD; Stefanie Stichling[1], MSc; Maxine Löhnert[1], MSc

Medicine with Specialization in Internal Medicine and General Medicine, Hochschule Fulda - University of Applied Sciences, Fulda, Germany

Corresponding Author:
Claudia Eberle, Prof Dr, MD
Medicine with Specialization in Internal Medicine and General Medicine
Hochschule Fulda - University of Applied Sciences
Leipziger Strasse 123
Fulda, 36037
Germany
Phone: 49 661 9640 ext 6328
Fax: 49 661 9640 648
Email: claudia.eberle@hs-fulda.de

Abstract

Background: The increasing prevalence of diabetes mellitus and associated morbidity worldwide justifies the need to create new approaches and strategies for diabetes therapy. Therefore, the ongoing digitalization offers novel opportunities in this field.

Objective: The aim of this study is to provide an updated overview of available technologies, possibilities, and novel insights into diabetes therapy 4.0.

Methods: A scoping review was carried out, and a literature search was performed using electronic databases (MEDLINE [PubMed], Cochrane Library, Embase, CINAHL, and Web of Science). The results were categorized according to the type of technology presented.

Results: Different types of technology (eg, glucose monitoring systems, insulin pens, insulin pumps, closed-loop systems, mobile health apps, telemedicine, and electronic medical records) may help to improve diabetes treatment. These improvements primarily affect glycemic control. However, they may also help in increasing the autonomy and quality of life of people who are diagnosed with diabetes mellitus.

Conclusions: Diabetes technologies have developed rapidly over the last few years and offer novel insights into diabetes therapy and a chance to improve and individualize diabetes treatment. Challenges that need to be addressed in the following years relate to data security, interoperability, and the development of standards.

(J Med Internet Res 2021;23(3):e23475) doi:10.2196/23475

KEYWORDS

diabetes mellitus; telemedicine; mobile apps; electronic health records; digital technology; eHealth; mobile phone

Introduction

Background

The ongoing digitalization changed the way of practicing medicine in general. Currently, it is possible to store and retrieve vast amounts of data and use different technologies to assist in diagnosis and therapy [1,2]. This technological development offers the opportunity to improve the treatment of diabetes mellitus (DM) in particular and in different ways [2,3].

In 2014, the global prevalence of DM among adults aged more than 18 years was estimated at approximately 8.5% and is believed to have increased since then, not taking into account the number of unreported cases [4]. As the high and rising prevalence and morbidity of DM, new strategies are needed.

In diabetes management, the measurement of blood glucose levels plays a key role. Therefore, different types of systems are available, which differ in when and how often the measurement takes place [5]. Furthermore, more than 200 million people with DM require insulin treatment [6]. Currently, there are mainly 2 possibilities to deliver insulin: multiple daily injections or continuous subcutaneous insulin infusions [5,6]. For both, different technological solutions are available in the market. Another important aspect of diabetes management depends on patient self-management. This means that people

with DM must make several decisions, for example, when and what to eat, when to exercise, and, if required, the insulin dose and timing [7]. Technological innovations such as mobile health (mHealth) apps or advances in telemedicine may be used in different ways to overcome challenges such as the ones mentioned above [2,8-17]. The purpose of diabetes technologies is to avoid complications of insulin therapy and to give people with DM more autonomy so that their quality of life (QoL) can be improved [18]. Furthermore, it must be noted that we have considered all types of diabetes and that they have different therapeutic needs. Type 1 and type 2 diabetes have different pathogeneses, leading to various therapy recommendations and needs. In addition to some similarities, type 1 diabetes therapy focuses more on insulin therapy, whereas type 2 diabetes management focuses more on nutrition and exercise. In a closer analysis of the technologies, it is important to differentiate between the types of diabetes and their needs.

Objective

The aim of this study is to provide an overview of the updated diabetes technologies available and the possibilities offered by digitalization of the treatment of DM.

Methods

Data Sources and Search Terms

We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for scoping reviews to assess the scope of the literature on the topic [19]. The electronic databases MEDLINE (PubMed), Cochrane Library, Embase, CINAHL, and Web of Science were searched. For example, in PubMed, the following search strategy was used for a title and abstract search: if available in MeSH terms: “Diabetes Mellitus” OR “diabetes” OR “blood glucose assessment” AND “Telemedicine” OR “Mobile Applications” OR “digitization” OR “digital treatment,” an additional search in the reference lists and Google Scholar was conducted.

Study Selection and Eligibility Criteria

The results of the literature search were transferred to the literature management program, Citavi, which was used to remove duplicates. Via a search of titles and abstracts, potentially relevant studies were extracted. Studies published in English or German between January 2008 and January 2020 that focused on technologies in DM therapy were included. Furthermore, to be included, the studies had to make remarks or comments on the clinical effectiveness, for example, effect on glycemic control, of the type of technology that was presented. There were no limitations to the study design. Studies that focused only on the comorbidities of DM were excluded. As this paper is a scoping review, we selected those studies from our presorted studies that seemed to give great overviews and explanations but still reported on clinical effectiveness.

Data Extraction and Data Synthesis

The included studies were thematically categorized according to the type of technology presented. We then extracted the following data from each study: author, year, study design, objective, and main findings.

Results

Overview

Our search yielded a total of 4935 results, of which we included 67 studies (including systematic and narrative reviews). The search results can be categorized into the following types of technology: glucose monitoring systems, insulin pens, insulin pumps, closed-loop systems, mHealth apps, telemedicine, and electronic medical records (EMRs).

Glucose Monitoring Systems

Most people with DM rely on self-monitoring of blood glucose (SMBG). Although there is a correlation between increased frequency of SMBG and a reduction in glycated hemoglobin A_{1c} (HbA_{1c}), SMBG only gives a short impression of blood glucose levels without any information about the change in glucose levels and glucose variability [5]. To gain this information, continuous glucose monitoring (CGM) systems can be used.

CGM systems measure the glucose level at regular intervals, 24 hours a day, and translate the measurements into data about glucose direction and rate of change [18]. On the basis of this information, CGM use without additional SMBG is as safe and effective as CGM use with SMBG support [20]. In addition, several experimental studies have shown that the use of CGM systems improves glycemic control through a significant reduction in HbA_{1c} [21-25], and this has been shown in combination with insulin pump therapy [22] and with multiple daily insulin injections [24]. Another positive effect was demonstrated by Jones et al [26] in their systematic review targeting pregnant women with pre-existing diabetes, showing that a reduction in hypertensive disorders is possible when CGM is used. However, they assumed only low-quality evidence because only 2 studies included the outcome [26]. Further aspects of CGM can improve diabetes self-management, flexibility, and QoL. In particular, the option for alarms and the possibility of device connection with smartphones or smartwatches give the participants a feeling of greater safety and freedom [27]. Furthermore, CMG measures subcutaneous glucose as a substitute for blood sugar. Old systems had to be calibrated regularly, whereas newer systems did not require calibration, which benefits the patients.

The Juvenile Diabetes Research Foundation recognized an association between greater CGM use and a greater decrease in HbA_{1c} levels. However, they also noticed a decrease in CGM usage over the study duration, especially in children and adolescents [21]. Reasons for discontinuation include discomfort when wearing the CGM system, problems with inserting the sensor, or with the adhesive holding the sensor on the skin [28]. An alternative to minimize skin reactions is to use implantable CGM systems, which can be as safe and accurate as transcutaneous CGM systems [29,30].

Most of the studies relate to the so-called real-time CGM. Another type of CGM system is the intermittently scanned CGM system, also called flash glucose monitoring. These terms refer to CGM systems that can only be used on demand. This means that patients must use a reader device to scan the sensor to

determine the glucose level [5]. This can be an alternative suitable for the masses because of lower costs relative to real-time CGM while also providing more information than with SMBG. However, it depends on the patients' willingness to scan the device several times a day [31,32].

Insulin Pens

Insulin pens are a technological invention for multiple daily insulin injections and are used as an alternative to insulin dosing via syringe. There are 2 types of insulin pens. Reusable pens have a delivery chamber in which insulin cartridges can be inserted. This makes it easy to change the type of insulin and is more economical than alternative prefilled pens. These contain a built-in single-use insulin cartridge. As a result, they are very convenient and easy to use, which is helpful for patients who have difficulties inserting the cartridges in reusable pens [33]. Overall, insulin pens have several advantages in terms of user friendliness, comfort of injections, and accuracy in delivering small doses of insulin. This leads to greater patient satisfaction and less reported pain [33,34].

However, there seems to be no improvement in glycemic control in insulin pens compared with the use of vials and syringes [35]. The advance in their development to so-called smart pens has changed this [36]. Smart pens include a memory function to track insulin doses and often have Bluetooth connectivity to transfer data to a smartphone app [18]. These functions seem to help patients improve their glycemic control [36].

The first FDA-approved smart pen, InPen, was launched in the United States in December 2017. In addition to digital diaries, it provides a dose calculator based on information on previous insulin doses, insulin on board, and current blood glucose levels. In addition, the app can be connected to some CGM systems [18,37,38]. This allows better diabetes management [39].

Insulin Pumps

Insulin pumps are small programmable computerized devices that allow continuous subcutaneous insulin infusions, so a more physiological insulin release can be mimicked. This is realized by an individual basal insulin rate, which is released hourly and set by the medical staff. In addition, bolus insulin can be released after meals or as a correction to reduce the target glucose levels [40,41]. The use of insulin pumps is associated with a reduction in HbA_{1c} levels without an increased risk of hypoglycemia and seems to be more effective than multiple daily insulin injections [42-44].

There are different types of insulin pumps [40], and in a patch insulin pump, the insulin needle is a part of the pump and is inserted when the pump is attached directly to the surface of the skin. Controls of these pumps are on the device itself, or a remote can be used, so that there is no tubing. Examples are the V-Go and Omnipod insulin pumps [45,46]. Tethered insulin pumps have tubing between the pump itself and the cannula, so the pump can be worn under or outside clothing or carried around in a pocket. Examples are the Medtronic MiniMed 670G and Tandem t:slim X2 insulin pumps [47,48]. Although they are rarely used, other types are implantable insulin pumps. These are implanted into the peritoneal cavity and remain there; therefore, the users have to travel to get the pump refilled with insulin [40].

The current development in insulin pump technology is a direct and wireless communication between insulin pumps and CGM systems, for example, via infrared technology. Such systems are called sensor-augmented insulin pumps (SAPs) [41]. In contrast to closed-loop systems, SAPs still require manual adjustment and input from the wearer [49,50]. However, the use of SAP can further reduce the risk of hypoglycemic episodes and improve glycemic control [51-53]. SAPs can have different features; however, they often include a bolus calculator, an automated insulin suspension, and custom reminders. The bolus calculator calculates insulin doses for meals after the number of carbohydrates the wearer intents to eat are put in [49]. An automated insulin suspension allows the SAP to suspend insulin delivery when the CGM detects an unexpectedly low glucose level or risk of hypoglycemia [51]. Custom reminders can be set to remind the individual to check their glucose levels or to set a bolus amount. In addition, safety alerts for missed bolus doses or blockage in the tubing or cannula can be set [49,54].

Closed-Loop Systems

Closed-loop systems, also referred to as artificial pancreas, are a combination of a CGM system and an insulin pump that can release insulin automatically [50]. These systems can be understood as fully automatic pumps that replace an endocrine pancreas; however, there are currently no such pumps in the market. Closed-loop pumps need to be calibrated, and patients need to assess their carbohydrate intake and tell the pump the time and amount of carbohydrates they will eat.

A step between SAPs and closed-loop systems are hybrid closed-loop systems such as MiniMed 670 G. This means that next to the manual mode where a basal insulin rate is set, they have the ability to use an automatic mode where insulin delivery is adjusted based on the CGM results. However, meal-related boluses remain user-dependent based on their carbohydrate intake [47,55,56]. Nevertheless, hybrid closed-loop systems have shown improved HbA_{1c} levels and time in range and decreased time spent in hypo- and hyperglycemias [55,57,58]. In addition, a closed-loop system consisting of the t:slim X2 insulin pump, Control-IQ Technology, and Dexcom G6 as a CGM system has been shown to improve glycemic control and time in range [59,60].

In addition, the #WeAreNotWaiting movement has developed systems for a do-it-yourself artificial pancreas (DIY AP). They used CGM systems, insulin pumps, and smartphone technology to run algorithms to calculate the insulin dose. These algorithms are openly shared by the community and have introduced various systems consisting of different components and using different interfaces. It must be noted that DIY AP are unregulated and therefore have potential safety concerns [61]. Although there are mainly observational studies about the effectiveness of DIY AP, they seem to improve the HbA_{1c} levels [62,63].

Telemedicine

Despite the advances in glucose monitoring systems and improvements in insulin delivery, the proportion of people with DM achieving their glycemic control targets has not improved

[64]. As a key contributor to poor glycemic control, *therapeutic inertia* is considered. *Therapeutic inertia* is defined as failure of initiation or intensification of therapy in a timely manner according to evidence-based clinical guidelines. It is a multifactorial problem, in which different stakeholders such as patients, health care providers, or the industry are involved [65,66]. Moreover, it underlines the need for care between visits to health care providers in diabetes treatment [67]. Telemedicine is becoming increasingly important as a solution to this problem [66]. Telemedicine describes the use of telecommunication systems to deliver health care at a distance, for example, through remote monitoring or real-time video conferencing [68]. This means that patients can communicate with their providers in an efficient way. However, health care providers can quickly provide feedback [67].

The open definition of telemedicine provides scope for various interventions, such as phone coaching interventions [69], tele-education [70,71], teleconsultations [72], real-time transmission of blood glucose levels [73], and web-based case management programs [74]. In addition, a combination of different components is possible [68,75]. Although there is a lack of evidence on long-term health outcomes, it is estimated that telemedical interventions lower HbA_{1c} levels with high certainty [68,76-79]. In addition, improvement in blood glucose and cholesterol, QoL, and the number of SMBG checks is possible [77,78]. This can be especially important in rural and medically underserved areas [71].

mHealth Apps

Apps for mobile devices such as smartphones focusing on health-related topics are called mHealth apps, and as the app market is very dynamic, they exist in a large variety. Diabetes apps vary in their range of considered topics and features [9,10,12,13,16,17,80-82]. Common features include tracking of blood glucose levels or insulin usage, calculating an insulin dose, monitoring of diet, weight or physical activity, providing education, or allowing communication and exchange of data with health care professionals or social networks [2,8-17,80-84].

Even with these features, there is great variability and development [16]. For example, attempts have been made to create bolus calculators that provide personalized and adaptive insulin recommendations [85] or to create apps that use augmented reality to improve estimations regarding carbohydrate intake [86]. With such features, apps support diabetes self-management, especially for people using SMBG [5,84,87-89]. In addition, some apps are able to connect with other devices such as CGM systems, smart pens, or insulin pumps. Some devices come with their own app, for example, apps belonging to CGM systems allow wireless synchronization between the glucometer itself and an individual's smartphone [81]. However, even with these features, not all needs expressed by people with DM are met. Adults with type 2 DM additionally wish for features that help in reducing the cognitive and emotional burden of diabetes self-management [90]. This is in line with recent findings showing that apps fail to provide relevant features to empower patients, even though there were complaints about a lack of personalized feedback, lack of education provided, and usability issues in the past [83,87,91]. Features such as writing a digital diary instead of a standard paper diary are not considered to affect QoL [92]. However, because of the combination of different features in most apps, it is difficult to determine the impact of specific app features on clinical outcomes [8]. Diabetes apps are supposed to improve glycemic control through significant reduction of HbA_{1c}, although evidence on safety and effectiveness is limited, especially regarding long-term outcomes [93-95].

Nevertheless, mHealth apps have several advantages. Nearly all patients and health care professionals have smartphones, and in contrast to diabetes devices, which are often left behind, smartphones are almost always in a patient's reach [81,96]. Moreover, apps have the ability to display data in colored charts and tables, which can be updated live and transferred via email or in cloud-based web portals [81]. In addition, an improvement in the reliability of patients' diabetes data is possible because errors such as illegible handwriting or incomplete data entries can be minimized [81,97]. However, some disadvantages should be considered before using or recommending apps. First, there is a risk that during constant updates, bugs emerge and features are removed without warning or new pricing models are introduced [81]. Furthermore, especially, older patients can have special needs, such as a large font size because of bad eyesight, so not all apps suit them [98]. Moreover, apps are mostly not regulated by a governing body unless they meet the definition of medical devices. This holds a potential safety risk, especially regarding possible misinformation and data security [2,81,94]. Finally, it must be considered that some apps are only available on one operating system, Android or iOS [81]. This means that individualized recommendations should be provided to people with DM.

EMR

The use of EMR offers the opportunity to transfer data of diabetes technology, such as glucose monitoring systems, insulin pumps, or insulin pens, directly to the patients' medical records. This makes everyday life in primary health care easier and is a step toward an automated institution. This can be realized by setting up stations where patients can transfer their data to their EMRs. To achieve this service for devices of different providers, collaboration with engineers and technicians is needed [99]. Another possibility is the use of apps to transfer data into the EMR. The prerequisite for this is that the systems used are compatible with one another, for example, the platform *HealthKit* of the mobile device company Apple can be used with compatible CGM systems and compatible EMR vendor patient apps [100].

In addition to their use in primary health care, the collected data of EMRs can be used in public health surveillance to better understand the disease [101]. However, at present, there is a lack of quality in the collected data and more structured variables are needed to exploit the full potential of this use [102].

Discussion

Principal Findings

In total, digital diabetology has a huge potential to improve the health and QoL of patients with DM, especially glycemic control

and diabetes therapy, using glucose monitoring systems, insulin pens, insulin pumps, closed-loop systems, mHealth apps, telemedicine, and EMR.

The variability of diabetes technologies shows that in the past decades, there has been a rapid development of care-supporting technologies. Advances in reducing the size and improving the accuracy of devices are not yet mentioned. It is believed that in the next 5 years, there will be many more advances, especially in CGM systems and automated insulin delivery through closed-loop systems. This includes the expectation that more systems will be available in the market [103]. Nevertheless, there are some points about the technologies that need to be discussed.

The advances in all the aforementioned technologies seem to improve glycemic control and, consequently, QoL. Therefore, they fulfill the purpose of diabetes technologies [18]. However, more data, especially on long-term outcomes, are needed [8,25,42,76,95]. Moreover, the technologies differ in the kind of improvement they make and their effectiveness. For example, insulin pens are an improvement compared with syringes in terms of pain and user friendliness; however, improvement in glycemic control seems to be possible with smart pens [33-36]. In contrast, insulin pumps are more likely to improve glycemic control than insulin pens [42-44]. However, insulin pump therapy does not suit everybody; for example, people doing athletics could experience problems with their pump because of failure or disconnection [39,104]. Therefore, an individualized solution that considers the patients' needs and expectations is needed. This applies, in particular, to the recommendation of a diabetes mHealth app. It is possible that younger patients are more likely to benefit from an mHealth app intervention or that older patients have special needs that must be considered [93,98]. In addition, factors such as patients' digital literacy should be considered. This is important because, despite the widespread use of smartphones, digital literacy barriers are common in vulnerable populations, which could reduce the effectiveness of diabetes technologies [105].

For such considerations, well-trained and educated health care personnel are needed [106]. Diabetes technologies can save time for patients who would spend less time at health care institutions and for health care personnel, because many processes can be automated [72,106]. However, this is still a type of future vision, because often enough missing interoperability between devices, especially of different providers, prohibits the automation of processes [107]. This goes hand in hand with missing standards and regulation of, for example, mHealth apps, which lead to potential health risks, for example, through misinformation [81,94,106]. In addition, the question of data safety is a major issue. Because these technologies deal with sensible patient data, protection must be of the highest priority. However, the missing regulation, specifically in mHealth apps, can lead to providers who store user data on remote servers that are more vulnerable to security breaches [81,94,106,107]. Therefore, there are still some aspects to address in the future.

Limitations

It should be noted that this study is a scoping review. This means that the presented data are not complete and provide an updated overview of the area of digital diabetology. Systematic reviews and meta-analyses should be taken into account to investigate the effects of technology type on diabetes therapy in more detail.

The technologies themselves also have limitations that must be considered. There is a time lag in tissue-glucose results and insulin effects, which affects digital therapy approaches. In addition, there must be a certain willingness on the part of the patient to use the technologies and the patient requires skills to use the tools.

Furthermore, some studies showed rather small intervention effects with regard to HbA_{1c} levels. The effect sizes must be examined in more detail.

Conclusions

In general, the different types of diabetes technologies offer a chance to individualize diabetology and improve the health and QoL of people with DM. With a look at the great advances that have been made in the last few years, rapid further development of diabetes technologies can be expected. In particular, the opportunity to connect different devices to create an automated insulin delivery will become more important. However, to use the full potential of digitalization, regulations and laws must be created to set standards and ensure data security.

Acknowledgments

This manuscript was created in the context of the project with the number EB 440/4-1 by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG). The authors would like to thank DFG for the support of this research.

Conflicts of Interest

None declared.

References

1. Cahn A, Akirov A, Raz I. Digital health technology and diabetes management. J Diabetes 2018 Jan;10(1):10-17. [doi: 10.1111/1753-0407.12606] [Medline: 28872765]
2. Eberle C, Stupin J, Schäfer-Graf U, Hummel M. Einsatz neuer technologien in diagnostik und therapie: diabetes-smartphone-apps. Diabetes in der Schwangerschaft: Praxisorientiertes Wissen zu Gestationsdiabetes, Diabetes mellitus Typ 1 und 2, MODY. 1. Auflage; 2020.

3. Miremberg H, Ben-Ari T, Betzer T, Raphaeli H, Gasnier R, Barda G, et al. The impact of a daily smartphone-based feedback system among women with gestational diabetes on compliance, glycemic control, satisfaction, and pregnancy outcome: a randomized controlled trial. Am J Obstet Gynecol 2018 Apr;218(4):453. [doi: 10.1016/j.ajog.2018.01.044] [Medline: 29425836]
4. Diabetes. World Health Organization. 2018. URL: https://www.who.int/news-room/fact-sheets/detail/diabetes [accessed 2021-02-13]
5. Akturk HK, Garg S. Technological advances shaping diabetes care. Current Opinion in Endocrinology & Diabetes and Obesity 2019;26(2):84-89. [doi: 10.1097/med.0000000000000467]
6. Garg SK, Rewers AH, Akturk HK. Ever-increasing insulin-requiring patients globally. Diabetes Technol Ther 2018 Jun;20(S2):21-24. [doi: 10.1089/dia.2018.0101] [Medline: 29873518]
7. Akturk HK, Rewers A, Joseph H, Schneider N, Garg SK. Possible ways to improve postprandial glucose control in Type 1 diabetes. Diabetes Technol Ther 2018 Jun;20(S2):224-232. [doi: 10.1089/dia.2018.0114] [Medline: 29916737]
8. Angelini S, Alicastro GM, Dionisi S, Di Muzio M. Structure and characteristics of diabetes self-management applications: a systematic review of the literature. Comput Inform Nurs 2019 Jul;37(7):340-348. [doi: 10.1097/CIN.0000000000000526] [Medline: 31136332]
9. Eberle C, Ament C. Individualisierte nutzung von diabetes-spezifischen mHealth-apps. Diabetologie und stoffwechsel 2015 Apr 29;10(S 01). [doi: 10.1055/s-0035-1549513]
10. Eberle C, Ament C. Digitale diabetologie – spezifische analyse von diabetes-apps hinsichtlich ihrer funktionen und nutzerbewertungen. Diabetologie und Stoffwechsel 2016 Apr 19;11(S 01). [doi: 10.1055/s-0036-1580864]
11. Eberle C, Ament C. Diabetes & Schwangerschaft 4.0 – Individualisierte Nutzungsanalyse von mHealth-Apps. Diabetologie und Stoffwechsel 2017 May 5;12(S 01):S1-S84. [doi: 10.1055/s-0037-1601701]
12. Eberle C, Ament C. Digitale diabetologie – update zur individuellen nutzungsanalyse diabetes-spezifischer mhealth-apps. Diabetologie und Stoffwechsel 2017 May 5;12(S 01):S1-S84. [doi: 10.1055/s-0037-1601700]
13. Eberle C, Ament C. Digitale diabetologie — Eine quantitative analyse diabetesspezifischer mHealth-apps. Diabetologie und Stoffwechsel (S01) 2019;14. [doi: 10.1055/s-0039-1688129]
14. Eberle C, Ament C. Schwangerschaft und digitalisierung — Individualisierte nutzungsanalyse von schwangerschafts- und GDM-apps. Diabetologie und Stoffwechsel 14 (S01) 2019. [doi: 10.1055/s-0039-1688270]
15. Eberle C, Ament C. Diabetes & Schwangerschaft – Individualisierte Nutzung von diabetes-spezifischen mHealth-Apps. Diabetologie und Stoffwechsel 2015 Apr 29;10(S 01). [doi: 10.1055/s-0035-1549513]
16. Eberle C, Ament C. Digitale diabetologie — Die "Epidemiologie" Diabetes-spezifischer mhealth-apps im zeitraum von 2015 bis 2018. Diabetologie und Stoffwechsel (S01) 2018;13. [doi: 10.1055/s-0038-1641863]
17. Eberle C, Ament C. Digitale diabetologie — Individuelle nutzungsanalyse diabetes-spezifischer mhealth-apps. Diabetologie und Stoffwechsel (S01) 2018;13. [doi: 10.1055/s-0038-1641949]
18. Acosta GJ, Brown S, Zand AM, Kansara A, Sadhu A. USA Vs Europe: who is leading the diabetes tech race? Curr Diab Rep 2019 Nov 16;19(11):128. [doi: 10.1007/s11892-019-1237-z] [Medline: 31734741]
19. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Sep 04;169(7):467-473. [doi: 10.7326/m18-0850]
20. Aleppo G, Ruedy KJ, Riddlesworth TD, Kruger DF, Peters AL, Hirsch I, REPLACE-BG Study Group. REPLACE-BG: a randomized trial comparing continuous glucose monitoring with and without routine blood glucose monitoring in adults with well-controlled type 1 diabetes. Diabetes Care 2017 Apr;40(4):538-545 [FREE Full text] [doi: 10.2337/dc16-2482] [Medline: 28209654]
21. Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. Effectiveness of continuous glucose monitoring in a clinical care environment: evidence from the Juvenile Diabetes Research Foundation continuous glucose monitoring (JDRF-CGM) trial. Diabetes Care 2010 Jan;33(1):17-22 [FREE Full text] [doi: 10.2337/dc09-1502] [Medline: 19837791]
22. Battelino T, Conget I, Olsen B, Schütz-Fuhrmann I, Hommel E, Hoogma R, SWITCH Study Group. The use and efficacy of continuous glucose monitoring in type 1 diabetes treated with insulin pump therapy: a randomised controlled trial. Diabetologia 2012 Dec;55(12):3155-3162 [FREE Full text] [doi: 10.1007/s00125-012-2708-9] [Medline: 22965294]
23. Hommel E, Olsen B, Battelino T, Conget I, Schütz-Fuhrmann I, Hoogma R, SWITCH Study Group. Impact of continuous glucose monitoring on quality of life, treatment satisfaction, and use of medical care resources: analyses from the SWITCH study. Acta Diabetol 2014 Oct;51(5):845-851 [FREE Full text] [doi: 10.1007/s00592-014-0598-7] [Medline: 25037251]
24. Beck RW, Riddlesworth T, Ruedy K, Ahmann A, Bergenstal R, Haller S, DIAMOND Study Group. Effect of continuous glucose monitoring on glycemic control in adults with type 1 diabetes using insulin injections: the DIAMOND randomized clinical trial. J Am Med Assoc 2017 Jan 24;317(4):371-378. [doi: 10.1001/jama.2016.19975] [Medline: 28118453]
25. Lind M, Polonsky W, Hirsch IB, Heise T, Bolinder J, Dahlqvist S, et al. Continuous glucose monitoring vs conventional therapy for glycemic control in adults with type 1 diabetes treated with multiple daily insulin injections: the GOLD randomized clinical trial. J Am Med Assoc 2017 Jan 24;317(4):379-387. [doi: 10.1001/jama.2016.19976] [Medline: 28118454]

26. Jones LV, Ray A, Moy FM, Buckley BS. Techniques of monitoring blood glucose during pregnancy for women with pre-existing diabetes. Cochrane Database Syst Rev 2019 May 23;5:CD009613 [FREE Full text] [doi: 10.1002/14651858.CD009613.pub4] [Medline: 31120549]
27. Ahmed Mohamed I, Fisher A, Cooper P, Hussain S. Use of continuous glucose monitoring in people with type 1 diabetes: perspectives of two people with diabetes and physician perspective. Diabetes Ther 2019 Apr 8;10(2):333-340 [FREE Full text] [doi: 10.1007/s13300-019-0576-8] [Medline: 30737675]
28. Wong JC, Foster NC, Maahs DM, Raghinaru D, Bergenstal RM, Ahmann AJ, T1D Exchange Clinic Network. Real-time continuous glucose monitoring among participants in the T1D Exchange clinic registry. Diabetes Care 2014 Oct;37(10):2702-2709 [FREE Full text] [doi: 10.2337/dc14-0303] [Medline: 25011947]
29. Kropff J, Choudhary P, Neupane S, Barnard K, Bain SC, Kapitza C, et al. Accuracy and longevity of an implantable continuous glucose sensor in the PRECISE study: a 180-day, prospective, multicenter, pivotal trial. Diabetes Care 2017 Jan;40(1):63-68. [doi: 10.2337/dc16-1525] [Medline: 27815290]
30. Christiansen MP, Klaff LJ, Brazg R, Chang AR, Levy CJ, Lam D, et al. A prospective multicenter evaluation of the accuracy of a novel implanted continuous glucose sensor: PRESCISE II. Diabetes Technol Ther 2018 Mar;20(3):197-206 [FREE Full text] [doi: 10.1089/dia.2017.0142] [Medline: 29381090]
31. Heinemann L, Freckmann G. CGM versus FGfM; or, continuous glucose monitoring is not flash glucose monitoring. J Diabetes Sci Technol 2015 Sep;9(5):947-950 [FREE Full text] [doi: 10.1177/1932296815603528] [Medline: 26330484]
32. Edelman SV, Argento NB, Pettus J, Hirsch IB. Clinical implications of real-time and intermittently scanned continuous glucose monitoring. Diabetes Care 2018 Nov;41(11):2265-2274. [doi: 10.2337/dc18-1150] [Medline: 30348844]
33. Baruah MP. Insulin pens: the modern delivery devices. J Assoc Physicians India 2011 Apr;59 Suppl:38-40. [Medline: 21823254]
34. Pearson TL. Practical aspects of insulin pen devices. J Diabetes Sci Technol 2010 May 01;4(3):522-531 [FREE Full text] [doi: 10.1177/193229681000400304] [Medline: 20513316]
35. Ebrahimi H, Pishgar F, Yoosefi M, Moradi S, Rezaei N, Djalalinia S, et al. Insulin pen use and diabetes treatment goals: a study from Iran STEPS 2016 survey. PLoS One 2019;14(8):e0221462 [FREE Full text] [doi: 10.1371/journal.pone.0221462] [Medline: 31461470]
36. Adolfsson P, Hartvig NV, Kaas A, Møller JB, Hellman J. Increased time in range and improved insulin adherence after introduction of a smart connected insulin pen. Diabetes Technol Ther 2020 Oct;22(10):709-718 [FREE Full text] [doi: 10.1089/dia.2019.0411]
37. Companion medical announces U.S. commercial launch of smart insulin pen system. Companion Medi. 2017. URL: https://www.prnewswire.com/news-releases/companion-medical-announces-us-commercial-launch-of-smart-insulin-pen-system-300571413.html [accessed 2021-02-13]
38. Klonoff DC, Kerr D. Smart pens will improve insulin therapy. J Diabetes Sci Technol 2018 May;12(3):551-553 [FREE Full text] [doi: 10.1177/1932296818759845] [Medline: 29411641]
39. Gildon BW. InPen smart insulin pen system: product review and user experience. Diabetes Spectr 2018 Nov 05;31(4):354-358 [FREE Full text] [doi: 10.2337/ds18-0011] [Medline: 30510392]
40. Sora ND, Shashpal F, Bond EA, Jenkins AJ. Insulin pumps: review of technological advancement in diabetes management. Am J Med Sci 2019 Nov;358(5):326-331. [doi: 10.1016/j.amjms.2019.08.008] [Medline: 31655714]
41. McAdams BH, Rizvi AA. An overview of insulin pumps and glucose sensors for the generalist. J Clin Med 2016 Jan 04;5(1) [FREE Full text] [doi: 10.3390/jcm5010005] [Medline: 26742082]
42. Misso ML, Egberts KJ, Page M, O'Connor D, Shaw J. Continuous subcutaneous insulin infusion (CSII) versus multiple insulin injections for type 1 diabetes mellitus. Cochrane Database Syst Rev 2010 Jan 20(1):CD005103. [doi: 10.1002/14651858.CD005103.pub2] [Medline: 20091571]
43. Reznik Y, Cohen O, Aronson R, Conget I, Runzis S, Castaneda J, et al. Insulin pump treatment compared with multiple daily injections for treatment of type 2 diabetes (OpT2mise): a randomised open-label controlled trial. The Lancet 2014 Oct;384(9950):1265-1272. [doi: 10.1016/s0140-6736(14)61037-0]
44. Calandro DA, Januszewski AS, Cuper KK, Burgess M, Horsburgh J, Loh M, et al. Substantial and sustained HbA1c reductions in australian insulin pump services for adults with type 1 diabetes. Benefit also evident for older and high HbA1c subjects. Madridge J Diabetes 2016 Oct 21;1(1):23-28. [doi: 10.18689/mjd-1000104]
45. Insulin delivery that lets you focus more on living your life. Valeritas. URL: https://www.go-vgo.com/how-v-go-works/using-v-go/ [accessed 2021-02-13]
46. Insulet Corporation. OmniPod insulin management system. Biomedi Safe and Stand 2015;45(20):155. [doi: 10.1097/01.bmsas.0000473415.86561.2b]
47. Medtronic. MiniMed 670G system user guide. MiniMed 670G System User Guide 2017 [FREE Full text]
48. t:slim X2 insulin pump. Tandem Diabetes Care. 2019. URL: https://www.tandemdiabetes.com/products/t-slim-x2-insulin-pump [accessed 2021-02-13]
49. Gilles G. How a Sensor-Augmented Insulin Pump (SAP) works. 2019. URL: https://www.verywellhealth.com/sensor-augmented-pump-sap-3289532 [accessed 2021-02-13]

50. Elleri D, Dunger DB, Hovorka R. Closed-loop insulin delivery for treatment of type 1 diabetes. BMC Med 2011 Nov 09;9:120 [FREE Full text] [doi: 10.1186/1741-7015-9-120] [Medline: 22071283]
51. Ly TT, Nicholas JA, Retterath A, Lim EM, Davis EA, Jones TW. Effect of sensor-augmented insulin pump therapy and automated insulin suspension vs standard insulin pump therapy on hypoglycemia in patients with type 1 diabetes: a randomized clinical trial. J Am Med Assoc 2013 Sep 25;310(12):1240-1247. [doi: 10.1001/jama.2013.277818] [Medline: 24065010]
52. Bergenstal RM, Tamborlane WV, Ahmann A, Buse JB, Dailey G, Davis SN, STAR 3 Study Group. Effectiveness of sensor-augmented insulin-pump therapy in type 1 diabetes. N Engl J Med 2010 Jul 22;363(4):311-320. [doi: 10.1056/NEJMoa1002853] [Medline: 20587585]
53. Steineck I, Ranjan A, Nørgaard K, Schmidt S. Sensor-augmented insulin pumps and hypoglycemia prevention in type 1 diabetes. J Diabetes Sci Technol 2017 Jan;11(1):50-58 [FREE Full text] [doi: 10.1177/1932296816672689] [Medline: 28264173]
54. Ly TT, Layne JE, Huyett LM, Nazzaro D, O'Connor JB. Novel bluetooth-enabled tubeless insulin pump: innovating pump therapy for patients in the digital age. J Diabetes Sci Technol 2019 Jan 21;13(1):20-26 [FREE Full text] [doi: 10.1177/1932296818798836] [Medline: 30239214]
55. Akturk HK, Giordano D, Champakanath A, Brackett S, Garg S, Snell-Bergeon J. Long-term real-life glycaemic outcomes with a hybrid closed-loop system compared with sensor-augmented pump therapy in patients with type 1 diabetes. Diabetes Obes Metab 2020 Apr;22(4):583-589. [doi: 10.1111/dom.13933] [Medline: 31789447]
56. Ekhlaspour L, Tabatabai I, Buckingham B. A review of continuous glucose monitoring data interpretation in the age of automated insulin delivery. J Diabetes Sci Technol 2019 Jul 26;13(4):645-663 [FREE Full text] [doi: 10.1177/1932296819851790] [Medline: 31130007]
57. Lal RA, Basina M, Maahs DM, Hood K, Buckingham B, Wilson DM. One year clinical experience of the first commercial hybrid closed-loop system. Diabetes Care 2019 Dec;42(12):2190-2196 [FREE Full text] [doi: 10.2337/dc19-0855] [Medline: 31548247]
58. Stone MP, Agrawal P, Chen X, Liu M, Shin J, Cordero TL, et al. Retrospective analysis of 3-month real-world glucose data after the minimed 670G system commercial launch. Diabetes Technol Ther 2018 Oct;20(10):689-692. [doi: 10.1089/dia.2018.0202] [Medline: 30160523]
59. Brown SA, Kovatchev BP, Raghinaru D, Lum JW, Buckingham BA, Kudva YC, iDCL Trial Research Group. Six-month randomized, multicenter trial of closed-loop control in type 1 diabetes. N Engl J Med 2019 Oct 31;381(18):1707-1717 [FREE Full text] [doi: 10.1056/NEJMoa1907863] [Medline: 31618560]
60. Brown S, Raghinaru D, Emory E, Kovatchev B. First look at Control-IQ: a new-generation automated insulin delivery system. Diabetes Care 2018 Dec;41(12):2634-2636 [FREE Full text] [doi: 10.2337/dc18-1249] [Medline: 30305346]
61. Crabtree TS, McLay A, Wilmot EG. DIY artificial pancreas systems: here to stay? Pract Diab 2019 Apr 17;36(2):63-68. [doi: 10.1002/pdi.2216]
62. Lewis D. Real-world use of open source artificial pancreas systems – poster presented at American Diabetes Association scientific sessions. OpenAPS. 2016. URL: https://openaps.org/2016/06/11/real-world-use-of-open-source-artificial-pancreas-systems-poster-presented-at-american-diabetes-association-scientific-sessions/ [accessed 2021-02-13]
63. Lewis DM, Swain RS, Donner TW. Improvements in A1C and Time-in-Range in DIY Closed-Loop (OpenAPS) users. Diabetes 2018 May;67(Supplement 1):352. [doi: 10.2337/db18-352-OR]
64. Carls G, Huynh J, Tuttle E, Yee J, Edelman SV. Achievement of glycated hemoglobin goals in the US remains unchanged through 2014. Diabetes Ther 2017 Aug;8(4):863-873 [FREE Full text] [doi: 10.1007/s13300-017-0280-5] [Medline: 28646411]
65. Khunti K, Gomes MB, Pocock S, Shestakova MV, Pintat S, Fenici P, et al. Therapeutic inertia in the treatment of hyperglycaemia in patients with type 2 diabetes: A systematic review. Diabetes Obes Metab 2018 Feb;20(2):427-437 [FREE Full text] [doi: 10.1111/dom.13088] [Medline: 28834075]
66. Garg SK, Parkin CG. The emerging role of telemedicine and mobile health technologies in improving diabetes care. Diabetes Technol Ther 2019 Jun;21(S2):21-23. [doi: 10.1089/dia.2019.0090] [Medline: 31169425]
67. Azar M, Gabbay R. Web-based management of diabetes through glucose uploads: has the time come for telemedicine? Diabetes Res Clin Pract 2009 Jan;83(1):9-17. [doi: 10.1016/j.diabres.2008.09.055] [Medline: 19056140]
68. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015 Sep 07(9):CD002098 [FREE Full text] [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]
69. Benson GA, Sidebottom A, Hayes J, Miedema MD, Boucher J, Vacquier M, et al. Impact of ENHANCED (diEtitiaNs Helping pAtieNts CarE for Diabetes) telemedicine randomized controlled trial on diabetes optimal care outcomes in patients with Type 2 diabetes. J Acad Nutr Diet 2019 Apr;119(4):585-598. [doi: 10.1016/j.jand.2018.11.013] [Medline: 30711463]
70. Buysse H, Coremans P, Pouwer F, Ruige J. Sustainable improvement of HbA and satisfaction with diabetes care after adding telemedicine in patients on adaptable insulin regimens: Results of the TeleDiabetes randomized controlled trial. Health Informatics J 2020 Mar;26(1):628-641 [FREE Full text] [doi: 10.1177/1460458219844369] [Medline: 31046527]

71. Kearns JW, Bowerman D, Kemmis K, Izquierdo RE, Wade M, Weinstock RS. Group diabetes education administered through telemedicine: tools used and lessons learned. Telemed J E Health 2012 Jun;18(5):347-353. [doi: 10.1089/tmj.2011.0165] [Medline: 22468984]
72. Charpentier G, Benhamou P, Dardari D, Clergeot A, Franc S, Schaepelynck-Belicar P, TeleDiab Study Group. The Diabeo software enabling individualized insulin dose adjustments combined with telemedicine support improves HbA1c in poorly controlled type 1 diabetic patients: a 6-month, randomized, open-label, parallel-group, multicenter trial (TeleDiab 1 Study). Diabetes Care 2011 Mar;34(3):533-539 [FREE Full text] [doi: 10.2337/dc10-1259] [Medline: 21266648]
73. Rodríguez-Idígoras MI, Sepúlveda-Muñoz J, Sánchez-Garrido-Escudero R, Martínez-González JL, Escolar-Castelló L, Paniagua-Gómez IM, et al. Telemedicine influence on the follow-up of type 2 diabetes patients. Diabetes Technol Ther 2009 Jul;11(7):431-437. [doi: 10.1089/dia.2008.0114] [Medline: 19580356]
74. McCarrier KP, Ralston JD, Hirsch IB, Lewis G, Martin DP, Zimmerman FJ, et al. Web-based collaborative care for type 1 diabetes: a pilot randomized trial. Diabetes Technol Ther 2009 Apr;11(4):211-217 [FREE Full text] [doi: 10.1089/dia.2008.0063] [Medline: 19344195]
75. Ralston JD, Hirsch IB, Hoath J, Mullen M, Cheadle A, Goldberg HI. Web-based collaborative care for type 2 diabetes: a pilot randomized trial. Diabetes Care 2009 Feb;32(2):234-239 [FREE Full text] [doi: 10.2337/dc08-1220] [Medline: 19017773]
76. Borries TM, Dunbar A, Bhukhen A, Rismany J, Kilham J, Feinn R, et al. The impact of telemedicine on patient self-management processes and clinical outcomes for patients with Types I or II Diabetes Mellitus in the United States: A scoping review. Diabetes Metab Syndr 2019;13(2):1353-1357. [doi: 10.1016/j.dsx.2019.02.014] [Medline: 31336491]
77. Rasmussen OW, Lauszus FF, Loekke M. Telemedicine compared with standard care in type 2 diabetes mellitus: A randomized trial in an outpatient clinic. J Telemed Telecare 2016 Sep;22(6):363-368. [doi: 10.1177/1357633X15608984] [Medline: 26468213]
78. González-Molero I, Domínguez-López M, Guerrero M, Carreira M, Caballero F, Rubio-Martín E, et al. Use of telemedicine in subjects with type 1 diabetes equipped with an insulin pump and real-time continuous glucose monitoring. J Telemed Telecare 2012 Sep;18(6):328-332. [doi: 10.1258/jtt.2012.120103] [Medline: 22912487]
79. Larsen ME, Turner J, Farmer A, Neil A, Tarassenko L. Telemedicine-supported insulin optimisation in primary care. J Telemed Telecare 2010;16(8):433-440. [doi: 10.1258/jtt.2010.100103] [Medline: 20841384]
80. Correia M. 47th ESCP symposium on clinical pharmacy Personalised pharmacy care. 24-26 October 2018, Belfast, Northern Ireland. Int J Clin Pharm 2019 Feb;41(1):289-383. [doi: 10.1007/s11096-018-0759-9] [Medline: 30488160]
81. Ahn DT, Stahl R. Is there an app for that? The pros and cons of diabetes smartphone apps and how to integrate them into clinical practice. Diabetes Spectr 2019 Aug 15;32(3):231-236 [FREE Full text] [doi: 10.2337/ds18-0101] [Medline: 31462879]
82. Goyal S, Morita P, Lewis GF, Yu C, Seto E, Cafazzo JA. The systematic design of a behavioural mobile health application for the self-management of type 2 diabetes. Can J Diabetes 2016 Feb;40(1):95-104. [doi: 10.1016/j.jcjd.2015.06.007] [Medline: 26455762]
83. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res 2011 Sep 22;13(3):e65 [FREE Full text] [doi: 10.2196/jmir.1874] [Medline: 21979293]
84. Eiland L, McLarney M, Thangavelu T, Drincic A. App-based insulin calculators: current and future state. Curr Diab Rep 2018 Oct 04;18(11):123. [doi: 10.1007/s11892-018-1097-y] [Medline: 30284645]
85. Pesl P, Herrero P, Reddy M, Xenou M, Oliver N, Johnston D, et al. An advanced bolus calculator for type 1 diabetes: system architecture and usability results. IEEE J Biomed Health Inform 2016 Jan;20(1):11-17. [doi: 10.1109/JBHI.2015.2464088] [Medline: 26259202]
86. Domhardt M, Tiefengrabner M, Dinic R, Fötschl U, Oostingh GJ, Stütz T, et al. Training of carbohydrate estimation for people with diabetes using mobile augmented reality. J Diabetes Sci Technol 2015 May;9(3):516-524 [FREE Full text] [doi: 10.1177/1932296815578880] [Medline: 25883165]
87. El-Gayar O, Timsina P, Nawar N, Eid W. Mobile applications for diabetes self-management: status and potential. J Diabetes Sci Technol 2013 Jan 01;7(1):247-262 [FREE Full text] [doi: 10.1177/193229681300700130] [Medline: 23439183]
88. Skrøvseth SO, Årsand E, Godtliebsen F, Joakimsen RM. Data-driven personalized feedback to patients with type 1 diabetes: a randomized trial. Diabetes Technol Ther 2015 Jul;17(7):482-489 [FREE Full text] [doi: 10.1089/dia.2014.0276] [Medline: 25751133]
89. Kim YJ, Rhee SY, Byun JK, Park SY, Hong SM, Chin SO, et al. A smartphone application significantly improved diabetes self-care activities with high user satisfaction. Diabetes Metab J 2015 Jun;39(3):207-217 [FREE Full text] [doi: 10.4093/dmj.2015.39.3.207] [Medline: 26124991]
90. Baptista S, Trawley S, Pouwer F, Oldenburg B, Wadley G, Speight J. What do adults with type 2 diabetes want from the "Perfect" app? Results from the second diabetes MILES: australia (MILES-2) study. Diabetes Technol Ther 2019 Jul;21(7):393-399. [doi: 10.1089/dia.2019.0086] [Medline: 31166804]

91. Brew-Sam N, Chib A. How do smart device apps for diabetes self-management correspond with theoretical indicators of empowerment? An analysis of app features. Int J Technol Assess Health Care 2019 Jan;35(2):150-159. [doi: 10.1017/S0266462319000163] [Medline: 31017563]
92. Drion I, Pameijer LR, van Dijk PR, Groenier KH, Kleefstra N, Bilo HJG. The effects of a mobile phone application on quality of life in patients with type 1 diabetes mellitus: a randomized controlled trial. J Diabetes Sci Technol 2015 May 11;9(5):1086-1091 [FREE Full text] [doi: 10.1177/1932296815585871] [Medline: 25963412]
93. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials. Diabetes Care 2016 Nov;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
94. Fleming GA, Petrie JR, Bergenstal RM, Holl RW, Peters AL, Heinemann L. Diabetes digital app technology: benefits, challenges, and recommendations. A consensus report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) Diabetes Technology Working Group. Diabetologia 2020 Feb;63(2):229-241. [doi: 10.1007/s00125-019-05034-1] [Medline: 31802144]
95. Veazie S, Winchell K, Gilbert J, Paynter R, Ivlev I, Eden KB, et al. Rapid evidence review of mobile applications for self-management of diabetes. J Gen Intern Med 2018 Jul;33(7):1167-1176 [FREE Full text] [doi: 10.1007/s11606-018-4410-1] [Medline: 29740786]
96. Smith A. Nearly half of American adults are smartphone owners. Pew Research Center. 2012. URL: https://www.pewresearch.org/internet/2012/03/01/nearly-half-of-american-adults-are-smartphone-owners/ [accessed 2021-02-13]
97. Mazze RS, Shamoon H, Pasmantier R, Lucido D, Murphy J, Hartmann K, et al. Reliability of blood glucose monitoring by patients with diabetes mellitus. Am J Med 1984 Aug;77(2):211-217. [doi: 10.1016/0002-9343(84)90693-4]
98. Isaković M, Sedlar U, Volk M, Bešter J. Usability pitfalls of diabetes mHealth apps for the elderly. J Diabetes Res 2016;2016:1604609 [FREE Full text] [doi: 10.1155/2016/1604609] [Medline: 27034957]
99. Wiesner T. Digitalisierung in der alltäglichen Praxis. Diabetologe 2018 Aug 13;14(7):455-459. [doi: 10.1007/s11428-018-0375-2]
100. Kumar RB, Goren ND, Stark DE, Wall DP, Longhurst CA. Automated integration of continuous glucose monitor data in the electronic health record using consumer technology. J Am Med Inform Assoc 2016 May;23(3):532-537 [FREE Full text] [doi: 10.1093/jamia/ocv206] [Medline: 27018263]
101. Allain TJ, Mang'anda G, Kasiya M, Khomani P, Banda NP, Gonani A, et al. Use of an electronic medical record to monitor efficacy of diabetes care in out-patients in a central hospital in Malawi: Patterns of glycaemic control and lessons learned. Malawi Med J 2017 Dec;29(4):322-326 [FREE Full text] [doi: 10.4314/mmj.v29i4.8] [Medline: 29963288]
102. Horth RZ, Wagstaff S, Jeppson T, Patel V, McClellan J, Bissonette N, et al. Use of electronic health records from a statewide health information exchange to support public health surveillance of diabetes and hypertension. BMC Public Health 2019 Aug 14;19(1):1106 [FREE Full text] [doi: 10.1186/s12889-019-7367-z] [Medline: 31412826]
103. Beck RW, Bergenstal RM, Laffel LM, Pickup JC. Advances in technology for management of type 1 diabetes. The Lancet 2019 Oct;394(10205):1265-1273. [doi: 10.1016/s0140-6736(19)31142-0]
104. Heinemann L, Fleming GA, Petrie JR, Holl RW, Bergenstal RM, Peters AL. Insulin pump risks and benefits: a clinical appraisal of pump safety standards, adverse event reporting and research needs. A joint statement of the European Association for the Study of Diabetes and the American Diabetes Association Diabetes Technology Working Group. Diabetologia 2015 May;58(5):862-870. [doi: 10.1007/s00125-015-3513-z] [Medline: 25784563]
105. Nouri SS, Avila-Garcia P, Cemballi AG, Sarkar U, Aguilera A, Lyles CR. Assessing mobile phone digital literacy and engagement in user-centered design in a diverse, safety-net population: mixed methods study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e14250 [FREE Full text] [doi: 10.2196/14250] [Medline: 31469083]
106. Heinemann L, Ickrath M. Digitalisierung: visionen für eine diabetologie der zukunft. Dtsch Arztebl 2017;114(41):27-28. [doi: 10.3238/PersDia.2017.10.13.07]
107. Mühlen H, Lueg A. Digitization in the doctor's office. Digitalisierungs- und Technologiereport Diabetes. 2019. URL: https://www.dut-report.de/2019/01/21/digitalisierung-in-der-arztpraxis/ [accessed 2021-02-13]

Abbreviations

CGM: continuous glucose monitoring
DIY AP: do-it-yourself artificial pancreas
DM: diabetes mellitus
EMR: electronic medical record
HbA_{1c}: glycated hemoglobin A_{1c}
mHealth: mobile health
QoL: quality of life
SAP: sensor-augmented insulin pump
SMBG: self-monitoring of blood glucose

Edited by G Eysenbach, R Kukafka; submitted 13.08.20; peer-reviewed by H Akturk, R Joakimsen; comments to author 03.09.20; revised version received 13.10.20; accepted 18.01.21; published 24.03.21.

Please cite as:
Eberle C, Stichling S, Löhnert M
Diabetology 4.0: Scoping Review of Novel Insights and Possibilities Offered by Digitalization
J Med Internet Res 2021;23(3):e23475
URL: https://www.jmir.org/2021/3/e23475
doi:10.2196/23475
PMID:33759789

Review

Human Coaching Methodologies for Automatic Electronic Coaching (eCoaching) as Behavioral Interventions With Information and Communication Technology: Systematic Review

Ayan Chatterjee[1], MEng; Martin Gerdes[1], PhD; Andreas Prinz[1], PhD; Santiago Martinez[2], PhD

[1]Department for Information and Communication Technologies, Centre for e-Health, University of Agder, Grimstad, Norway

[2]Department of Health and Nursing Science, Centre for e-Health, University of Agder, Grimstad, Norway

Corresponding Author:
Ayan Chatterjee, MEng
Department for Information and Communication Technologies
Centre for e-Health
University of Agder
Jon Lilletuns Vei 9
Grimstad, 4879
Norway
Phone: 47 38141000
Email: ayan.chatterjee@uia.no

Abstract

Background: We systematically reviewed the literature on human coaching to identify different coaching *processes* as behavioral interventions and *methods* within those processes. We then reviewed how those identified coaching processes and the used methods can be utilized to improve an electronic coaching (eCoaching) process for the promotion of a healthy lifestyle with the support of information and communication technology (ICT).

Objective: This study aimed to identify coaching and eCoaching processes as behavioral interventions and the methods behind these processes. Here, we mainly looked at processes (and corresponding models that describe coaching as certain processes) and the methods that were used within the different processes. Several methods will be part of multiple processes. Certain processes (or the corresponding models) will be applicable for both human coaching and eCoaching.

Methods: We performed a systematic literature review to search the scientific databases EBSCOhost, Scopus, ACM, Nature, SpringerLink, IEEE Xplore, MDPI, Google Scholar, and PubMed for publications that included personal coaching (from 2000 to 2019) and persuasive eCoaching as behavioral interventions for a healthy lifestyle (from 2014 to 2019). The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework was used for the evidence-based systematic review and meta-analysis.

Results: The systematic search resulted in 79 publications, including 72 papers and seven books. Of these, 53 were related to behavioral interventions by eCoaching and the remaining 26 were related to human coaching. The most utilized persuasive eCoaching methods were personalization (n=19), interaction and cocreation (n=17), technology adoption for behavior change (n= 17), goal setting and evaluation (n=16), persuasion (n=15), automation (n=14), and lifestyle change (n=14). The most relevant methods for human coaching were behavior (n=23), methodology (n=10), psychology (n=9), and mentoring (n=6). Here, “n” signifies the total number of articles where the respective method was identified. In this study, we focused on different coaching methods to understand the psychology, behavioral science, coaching philosophy, and essential coaching processes for effective coaching. We have discussed how we can integrate the obtained knowledge into the eCoaching process for healthy lifestyle management using ICT. We identified that knowledge, coaching skills, observation, interaction, ethics, trust, efficacy study, coaching experience, pragmatism, intervention, goal setting, and evaluation of coaching processes are relevant for eCoaching.

Conclusions: This systematic literature review selected processes, associated methods, strengths, and limitations for behavioral interventions from established coaching models. The identified methods of coaching point toward integrating human psychology in eCoaching to develop effective intervention plans for healthy lifestyle management and overcome the existing limitations of human coaching.

(J Med Internet Res 2021;23(3):e23533) doi:10.2196/23533

KEYWORDS

coaching; electronic coaching; human behavior; healthy lifestyle; persuasive technology

Introduction

Overview

A coach [1-4] is a trusted role model, adviser, wise person, friend, mensch (a person of integrity and honor), steward (supervisor), or guide. A coach facilitates experimental learning that results in future-oriented abilities. Coaches can shape new visions and plans to achieve desired results. Coaching has been implemented in management, leadership, entrepreneurship, health care, and performance management. It helps participants to cultivate themselves and become more successful in achieving their set goals. Successful coaching relies on a good relationship, mutual trust, and freedom of expression between coaches and participants [1-6]. Effective coaching leads to excellent performance, self-motivation, and self-correction. Coaching processes can be divided into the following two categories: (1) traditional offline human coaching (coaching by humans) and (2) electronic coaching (eCoaching).

Traditional offline human coaching processes involve the following methods [1,4,7-10]: privacy, focus, goal orientation, performance improvement, and trust. The process associated with coaching by humans can be achieved either face-to-face or remotely (via telematic means). Furthermore, the coaching process can be categorized [5-16] as health coaching to address negative behavioral change, cognitive-behavioral coaching, mental health coaching, in-house executive coaching in businesses, companies, or industries (corporate coaching), sports coaching, motivational coaching, educational coaching, and coaching to carry out activities of daily living. Traditional human coaching is a dialogic, goal-oriented, pragmatic learning practice. The human coaching process can be further enhanced through electronic modes, such as video, audio, email, chatbot, and text, with the support of information and communication technology (ICT), which is referred to as eCoaching. In the last decade, personal coaching for behavioral intervention has been increasingly used to promote a healthy lifestyle [17,18]. eHealth uses ICT for health [19,20]. eCoaching is a promising eHealth research direction for continuous customized ways of lifestyle support [21,22]. It is an evolution of offline human coaching with the flexibility of electronic services allowing ubiquitous access to the process. eCoaching technologies represent an evolving trend in the domain of human behavioral intervention. The coaching core behind an eCoaching system can be a human (eg, telemedicine), an artificial intelligence (AI) agent (eg, algorithm), or a combination of these. An eCoaching system consists of a set of programmed modules representing an artificial entity that may look at, query, examine from, and predict a consumer's behaviors in a specific context and in a specific period. Application domains of eCoaching include the following [18,23-55]: nutrition coaching, physical activity coaching, coaching for mental health, coaching for activities of daily living in the elderly, diabetic coaching, and cardiac rehabilitation. Studies in eCoaching can offer methods to enhance individual healthcare with ICT. A virtual eCoaching recommendation system can guide people and convey the appropriate recommendations in real time to improve their lifestyle [21,22,56]. The leading methods of eCoaching processes are monitoring, decision making, goal setting, persuasion, awareness provision (intervention), goal evaluation, and learning for future actions [24,27,32-34,57-59]. Digital techniques of lifestyle change with eCoaching have appeared as efficient and scalable options for intensive behavioral counseling when face-to-face or in-person programs are inaccessible or undesirable. eCoaching can make human behavioral interventions useful when combined with human coaching methods [57,60,61]. In the eCoaching processes, participants can remotely take part and avoid traveling, expenses, and transport risks. It is relevant to note that eCoaching will electronically handle data. Therefore, complying with general data protection regulations is critical for the safety and security of participants. eCoaching processes may ideally influence health outcomes, for which aspects, such as usability, efficacy, and adherence, may play important roles to influence health and/or health behavior. "Efficacy" means the effects of behavioral intervention following any coaching process (of any method, not only of eCoaching). "Usability" means the effectiveness, efficiency, and satisfaction when using a technology. "Adherence" means the degree to which the technology is used as intended [7,57,58].

Coaching as a human behavioral intervention is a personalized planned process designed to reward and reinforce the positive behavior of human beings. Each behavioral intervention differs from others based on the participants who are the primary targets of the intervention, where psychology and context play crucial roles [21,56,60]. The methods of a successful behavioral intervention plan "focus" on the identification of problems, the analysis of identified problems, prevention strategies and modification techniques, encouragement or motivation, strategic planning to diminish negative behavior, and participant engagement [1-3,5,35,56,60,62]. The coaching process for behavioral intervention should include appropriate guidelines, mutual trust, a rewarding plan, participant feedback, goal setting, and goal evaluation methods to make it useful for coaching and eCoaching (coaching by an electronic coach [eCoach]) processes [1,23,63]. Time is a critical factor in determining the format of coaching. Integration of coaching methodologies into persuasive eCoaching for electronic personalized behavioral interventions creates new opportunities for a healthy lifestyle [1-3]. It is rewarding for participants to change negative behavior using evidence-based methods and to observe the increase in their health and strength [4,5,60].

Aim of the Study

The aim of this systematic literature review was to identify key processes from the coaching methodologies to tackle the existing challenges coupled with the human coaching and eCoaching processes as behavioral interventions. The focus on coaching is justified by the fact that health and wellness remote coaches are an asset to clinical practice although they are underutilized in the health care system [6,21,56].

This systematic literature review addresses the following research questions (RQs):

(1) RQ1: What are the existing human coaching processes?

(2) RQ2: Which conceptual coaching models can be used to explain the coaching process?

(3) RQ3: What are the basic coaching methods to make coaching processes successful for the promotion of a behavioral intervention?

(4) RQ4: How can the methods of human coaching processes be incorporated into eCoaching for behavioral intervention to promote a healthy lifestyle?

(5) RQ5: How can eCoaching promote a healthy lifestyle with proven coaching methods using ICT?

Methods

A systematic literature review was used to acquire a comprehensive overview of the current literature on the topic in a reproducible and transparent way. Systematic reviews represent a scientific synthesis [64,65] of evidence. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) evidence-based framework [64] was used for the systematic review and meta-analyses. Initially, we performed a random search in the "Google Scholar" database with the following four key terms: "coaching," "electronic coaching," "eCoaching," and "e-Coaching" (see Table 1 for the results of the initial random search). It was observed that the keyword "electronic coaching" obtained the greatest number of results among the last three key terms.

Table 1. Initial "Google Scholar" random literature search results according to publication year.

Key terms	1998-2019	2008-2019	2014-2019	2017-2019
Coaching, n	482,000	497,000	159,000	50,800
Electronic coaching, n	133,000	72,600	25,700	18,400
eCoaching, n	396	347	270	184
e-Coaching, n	5740	5100	3300	1710

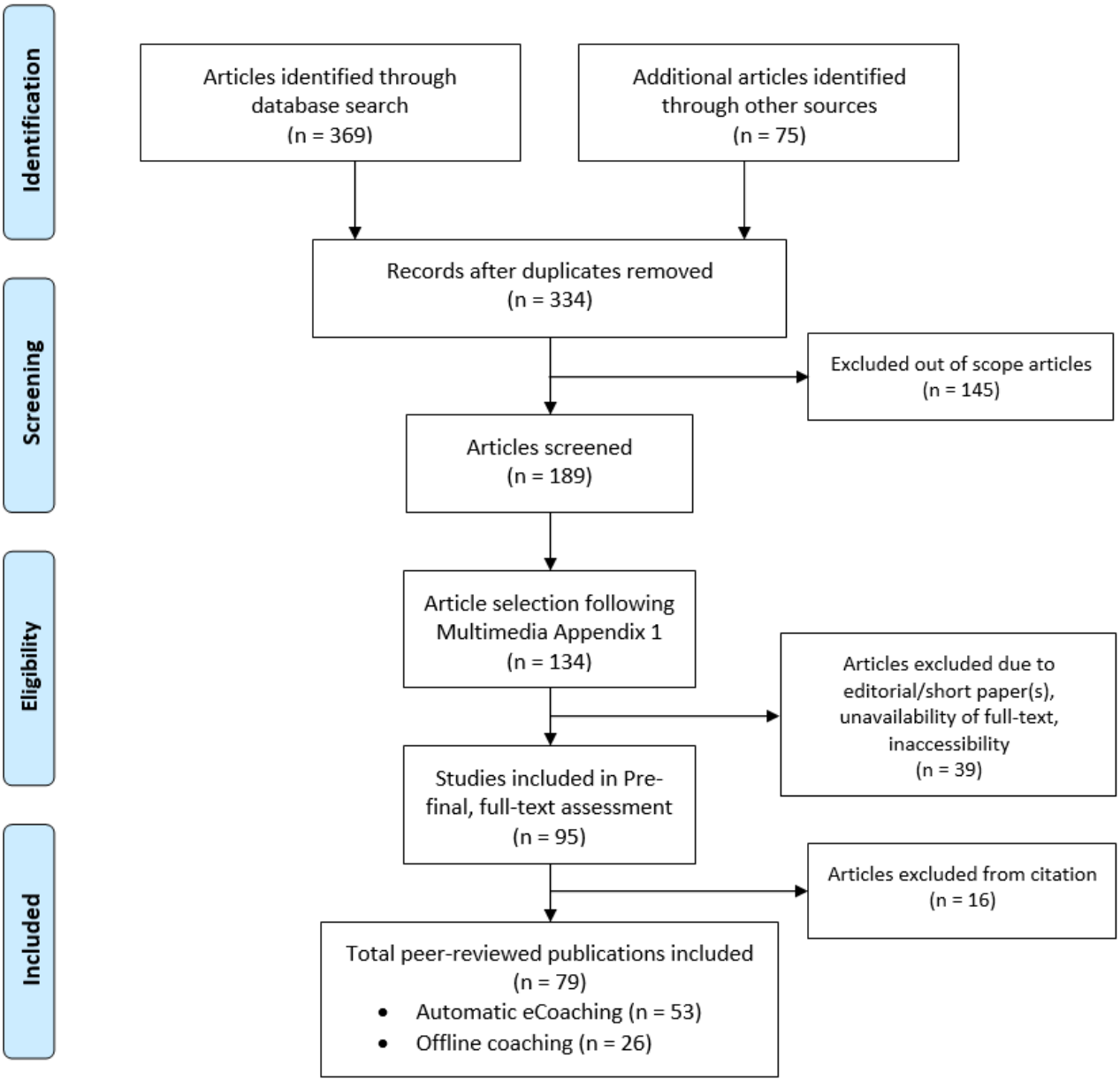
Subsequently, literature searches were conducted with selected search string patterns (Table 2) on the following electronic databases, as they compiled the greatest number of scientific sources related to coaching and eCoaching studies: Google Scholar, EBSCOhost, Scopus, ACM, Nature, SpringerLink, IEEE Xplore, MDPI, and PubMed. This study's search strategy was created in collaboration with the library of the University of Agder (UiA) in Norway, based on the following two main search topics: (1) coaching as a behavioral intervention and (2) eCoaching as a behavioral intervention. Related search keywords were identified using MeSH (Medical Subject Headings) terms, synonyms, keywords from relevant articles, and self-determined search terms. The means, such as EndNote (V. X9), DOAJ, Sherpa/Romeo, and Microsoft Excel (MS Office 365 V. 16.x), were used to effectively search, collect, and select related articles. We aimed to include articles that described coaching methodologies and eCoaching related to behavioral interventions. Articles were categorized among the groups quantitative, qualitative, and editorial. The quantitative study deals with statistical analysis on systematically collected data to test a specific hypothesis, while the qualitative study focuses on words and meanings to explore ideas and experiences in depth. The search results are depicted in Multimedia Appendix 1. We included articles based on the following inclusion criteria: (1) peer-reviewed, full length articles written in English, (2) eCoach articles published in the selected databases between 2014 and 2019, (3) coaching articles published in the selected databases between 2000 and 2019, (4) articles indexed in "Google Scholar," (5) journal papers, conference papers, or books, (6) qualitative (primary and secondary research) and quantitative studies, and (7) coaching articles related to human behavioral intervention. The traditional offline human coaching processes are older than eCoaching processes. Thus, the period for searching the selected electronic databases differs for "coaching" and "eCoaching."

Table 2. Search strings used for article searching.

Category	Search strings	Publication year
eCoach	(mentoring OR "e-coach" OR "ecoach" OR "electronic coach*" OR counseling OR educat* OR electro*coach*) OR (telemedic* OR "mobile health" OR mhealth OR ehealth OR "e-therap*" OR "e-counseling") AND (obesity OR overweight OR overnutrit* OR hypernutrit* OR lifestyle OR behavior OR behaviour) AND (persuasion OR recommendation OR intervention)	2014-2019
Coach	(mentoring OR coaching OR counseling OR educat* OR coach* OR executive* OR sport* OR activity* OR life*) AND (health* OR behavior OR behaviour OR psychology OR lifestyle)	2000-2019

We excluded editorial articles, studies related to robotic coaching, philosophical papers, articles with a lot of similar content or articles that were exactly repeated, and articles that were neither "open access" nor accessible through the university library. The full process of selecting sources for this review is depicted in a flowchart (Figure 1). The process includes the following four phases [64]: identification, screening, eligibility, and inclusion.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for the article selection process.



Results

Literature Search Results

The searches (electronic database and manual searches) resulted in 444 papers (369 in electronic databases and 75 identified manually), where 110 were duplicates. In the prefinal stage, we selected 95 articles for full-text review after checking the abstract, conclusion, length of the paper, and availability of full text. In the final search, we included peer-reviewed publications only, resulting in 79 core peer-reviewed articles related to "coaching" (53 papers) and "eCoaching" (26 papers). The categorical distribution (quantitative/qualitative) of the selected articles under "eCoaching" and "coaching" was as follows: coaching (23 quantitative, 30 qualitative) and eCoaching (7 quantitative, 19 qualitative).

This systematic literature review identified different coaching process descriptive models, as well as how they are carried out and in which context. We observed underlying theories to support traditional human coaching processes, such as hope theory [7] and amoeba theory [2], and different terms associated with both coaching and eCoaching processes, such as components, conceptual models, aspects, principles, concepts, activities, and methods. The usage of heterogenous terms to describe similar or nearly similar coaching and eCoaching processes resulted in ambiguity, less contentedness, and reduced clarity. Therefore, throughout the study, we concentrated on *processes* and *methods* to answer and discuss our research questions. They can be explained as follows: *processes* describe different coaching and eCoaching models and their implementation style, and the success of a coaching and eCoaching process depends on the adopted *methods*. Their evaluation helps to determine the performance of the human coaching and eCoaching processes.

The identified methods help us to understand the principles, strategies, effectiveness, and constraints of coaching and

eCoaching processes. An eCoach may create optimized, real-time, comprehensible, automated, contextual, evidence-based, and personalized intervention strategies for participants. Moreover, an eCoach may address the challenges associated with coaching, such as scope, the volume of the target audience, bias, cost, automation, accessibility, security, flexibility, credibility, conceptual clarity, location, and time independence, as revealed from the systematic literature review [2,4,6,21,35,56,62,66-70]. This systematic literature review identified 21 studies contributing to answering RQ1 regarding coaching methodologies; 17 studies contributing to answering RQ2 regarding a conceptual coaching model; 20 studies contributing to answering RQ3 regarding coaching methods for the promotion of "behavioral intervention;" 59 studies contributing to answering RQ4 regarding the integration of "coaching process" into "eCoaching for behavioral intervention;" and 35 studies contributing to answering RQ5 to advance "eCoaching for behavioral intervention" for a "healthy lifestyle" with proven "coaching methodologies" using ICT (several included overlapped studies contribute to multiple RQs).

RQ1: What Are the Existing Human Coaching Processes?

Bartlett [1] proposed a method where mutual trust, respect, and freedom of expression were considered as the elements of a successful coaching relationship. The model combines the establishment of a relationship between a coach and a trainee, recognizing an opening to assess obstacles related to coaching, observation, and assessment; enrollment of clients; and coaching conversations. Potrac et al [66] proposed another model that combines systematic observation and interpretive interview techniques to gain a deeper and broader understanding of personal coaching's instructional process. The suggested multimethod framework concerns identification of the instructional behaviors within the practice environment, generation of the understanding of why coaches behave as they do within the practice environment, and examination of the impact on the instructional strategies and understanding by humans. Cunningham et al [9] recommended a model based on hierarchical regression analysis with a stepwise process to show that an earlier success accuracy, collective coaching experience, collective professional coaching experience, and racial diversity are significantly associated with team performance. Côté [5] proposed that informal self-directed learning modes have relatively more significance than formal and nonformal learning. The proposed model combines the following three variables: (1) individuals with different backgrounds, experiences, and knowledge, (2) coaching work in various types of contexts with varying amounts of resources, equipment, and facilities, and (3) coaching work with participants varying in terms of age, developmental level, and goals. Their proposed coaching model divides variables into the two categories of ambient components (such as coach's and participant's characteristics, and contextual factors) and behavioral components (such as competition, organization, and training). The model proposed by Green et al [7] included the concept of coaching psychology and hope theory, based on the belief that human actions are goal directed. They claimed that the cognitive-behavioral solution-focused coaching model provides preliminary evidence on life coaching that can enhance mental health, quality of life, and goal attainment. Goal setting and goal evaluation are central to lifestyle coaching and are the pillars of successful self-regulation. The coaching study focused on evaluating the effectiveness of a cognitive-behavioral, solution-focused, life coaching group program, and its impact on goal striving, well-being, and hope. The assessment included measures of the "Satisfaction with Life Scale (SWLS)" and the "Positive and Negative Affect Scale (PANAS)." Murphy et al [6] proposed a model focused on executive coaching. With the support of conceptual clarity, executive coaching could unify efforts and resources and provide a common understanding to enhance human resource developmental programs. The human resource should play an active role in developing the organizational capacity for leadership. The proposed model of Richards [67] combines a recurrent process of suitable environment creation for coaching, learning for innovation and successful adaptation, and achievement (coaching performance) for sustained performance. Coaches need to rethink the discipline of coaching as if it is performed well, and coaching can increase motivation and contribute to sustaining high performance. Another model proposed by Richards [67] combines "tell" and "do" instructions. The model is based on the method of conventional thinking to improve a participant's performance by telling them or showing them what they are doing wrong in order to avoid any repetitive mistakes. This model is beneficial within a short-term context and frame of mind. However, overuse of the approach will undermine efforts to achieve long-term performance. The proposed model by Flaherty [2] was drawn from the concept of phenomenology and combines the following five methods in the coaching process: relationship building (based on mutual satisfaction, mutual respect, mutual trust, and freedom of expression), pragmatism (persistent correction following a feedback loop), two tracks (client and coach engagement), always/already (intervention planning), and identification of techniques that do not work (identification of challenges/limitations). The proposed amoeba theory is discussed based on behaviorism and is used in management for changing behavior either by poking or giving rewards. Cox [3] proposed a model that is based on adult learning and human psychology. The study included the following eight learning theories relevant to coaching: andragogy, transformative learning, reflective practice, experimental learning, learning styles, life course development, values and motivation, and self-efficacy. The proposed model by Stober et al [4] focusses on a humanistic approach to the process of coaching with four guiding principles, including the nature of the coaching relationship, the client as a source and director of change, the client as a whole and distinct person, and the coach as the facilitator of the client's growth. The model proposed by Knight [68] includes the method of instructional coaching. Visible learning (diagnosis, intervention, and evaluation) has been one of the research initiatives conducted in education in the past few decades. Simultaneously, instructional coaching (identify, learn, and improve) is becoming a popular form of professional development. Instructional coaching is used to support the realization of "visible learning" or other educational innovations. Standing [16] proposed a model to compare the use of "traditional" and "progressive"

coaching styles to train a general male youth population to improve sprint and jump performances while assessing enjoyment in order to comment on the long-term application. The process includes the following steps: study design, participant selection, experimental procedure, data collection, statistical analysis, and performance measurement.

RQ2: Which Conceptual Coaching Models Can Be Used to Explain the Coaching Process?

The actual definition of coaching concepts remains difficult to understand, and the working of the coaching interaction itself is still unknown [3]. The coaching approach may create a positive impact on the coaching environment and, subsequently, can improve the bottom-line performance of a target human group. An efficient coaching model is a tool to motivate personal learning, increase energy, improve ownership, and improve accountability. In contrast, no single coaching model can be labeled as the best, as coaching models change with the perspective and context of individual coaching. We found coaching model candidates for behavioral intervention that adequately explain the human coaching process. We divided our findings into the following two categories to have a better understanding of coaching process descriptive models: coaching process descriptive models and their application domain (Table 3 [3,7,8,71]; Figure 2 [8], Figure 3 [71], and Figure 4 [3]), and psychological approaches to describe coaching process models (Table 4 [4,66,67,72]).

Table 3. Coaching process descriptive models and their application domain.

Coaching process descriptive models	Application domain
Five level reviewing belief model [8]: The model explains the learned belief of a client who inherits or learns beliefs from his/her ancestors (parents) or teachers, or influential people to take any action from a decision. This model has the following five different levels from the bottom to the top: review (who am I), define (where do I want to go), plan (how am I going to get there), identify (how do I need to think, feel, behave), and continue (review and reward). When action is taken after a decision is made, the following five different levels are explored: beliefs and values, thoughts and expectations, emotions, behaviors, and actions. The model is shown in Figure 2. *ABCDE model* [8]: It explains how to use the tools and techniques of cognitive behavioral coaching to challenge negative thinking, make positive changes, achieve goals, and improve (ABCDE model: A, activating event or situation; B, the belief; C, the consequential emotion; D, disputing the belief; E, exchanging the thought). *Cognitive behavioral model* [7]: It utilizes a cognitive behavioral solution-focused model of coaching. It provides preliminary evidence that evidence-based life coaching can enhance mental health, quality of life, and goal attainment.	Cognitive behavioral coaching
Dynamic coaching model [71]: A coaching system is made up primarily of three spaces that contain three conversations that interact together to create the coaching conversation. The first reflective space is the internal conversation within the client. The second space is the shared space created in between the coach and client. The third space is the space within the coach. The complete coaching model is depicted in Figure 3.	Dynamic coaching
Experiential coaching cycle [3]: It has the following three noticeable constituent areas: prereflective experience, reflection on experience, and postreflective thinking, as depicted in Figure 4. The cycle additionally has the following three essential transition stages: touching experience, turning into critical, and integration. Transition phases regularly involve more emotional, cognitive, or physical effort than the constituent spaces and are particularly challenging for both coach and participant owing to the emotional struggle and inheritance of uncertainty.	Pragmatic coaching

Figure 2. The five-level reviewing belief model by Whitten.

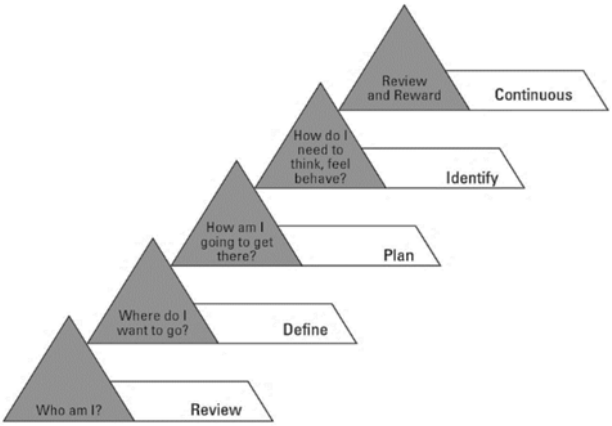


Figure 3. The dynamic coaching model by Cavanagh et al.

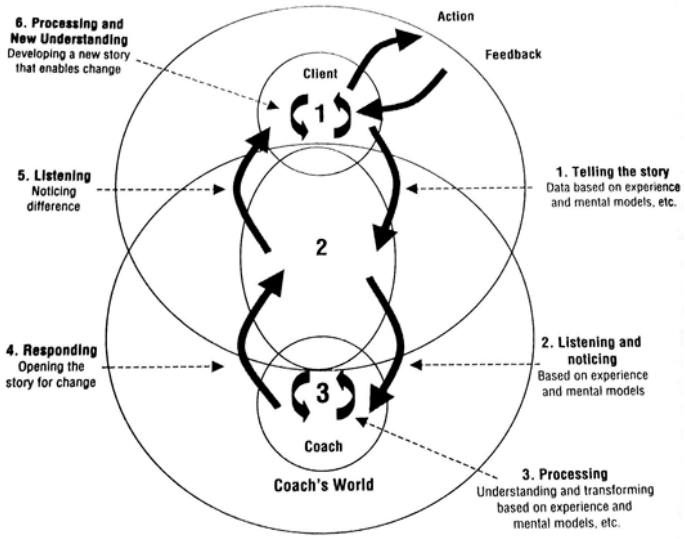


Figure 4. The experiential coaching cycle with six phases by Cox.

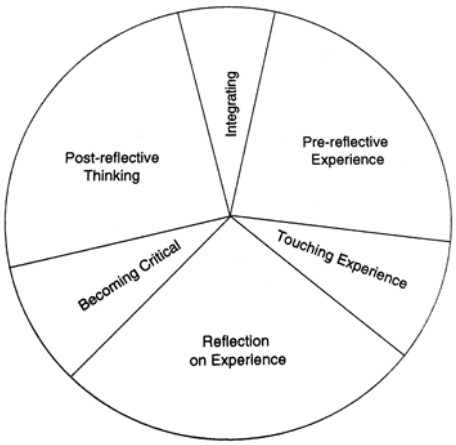


Table 4. Psychological approaches for coaching process models.

Coaching process descriptive models	Psychological approaches
Five elements model [66]: The model explains the practice for human resource development, focusing on improving performance/examining results with a way of equipping human beings with the methods, knowledge, and possibilities they want to broaden themselves and become more effective. The unidirectional sequence of five elements includes establishing relationships, recognizing opening, observation or assessment, enrollment of clients or participants, and coaching conversations. *Goal focused executive coaching model* [4]: It explains how to improve personal or professional performance, personal satisfaction, and effectiveness in the client's organization within a formally defined coaching agreement, and identify a set of goals using the following process: identification of an issue, setting of a goal, development of a cyclic action plan (act, monitor, evaluate, and change), and evaluation of the success score. *Organization response cycle* [67]: The model explains how to manage the pressure exerted on the department because of globalization, in order to produce faster, cheaper, and customized products and services. This model includes a cyclic loop of the following processes: learning (individual, team, organizational), innovation (products, and services), adaptation (responding to change and complexity), and results (enough or not).	Executive coaching
Humanistic coaching model [4]: The cyclic model of awareness-choice-execution (ACE) explains how to use the principles and tasks to teach participants how to harness their own growth process. In directing the process of coaching for change, the coach can ensure that the participant integrates "being (and awareness of that)" with "doing" such that the participant comes away with actual results.	Humanistic coaching
Effective coaching model [4]: It explains the core of the coaching process ("what is done!") and represents how the contextual themes are legislated with the following seven key principles that strengthen the human coaching process: collaboration, accountability, awareness, responsibility, commitment, action, and results.	Contextual coaching
Open innovation model [72]: It explains several key factors for organization development throughout the following life cycle stages: birth (innovation, awareness, intuition, vision, commitment, risk, and flexibility), growth (decision making, delegation, team approach, state change, and ability to grow), maturity (feasibility, retain high performance, overcome obstacles, and responsiveness), revival (autonomy, integration, effective internal communication, and innovative high performance), and decline (renew strategy and structure, innovativeness, improve information processing, and increase tolerance level).	Organization development

RQ3: What Are the Basic Coaching Methods to Make the Coaching Process Successful for the Promotion of a Behavioral Intervention?

A coach must sometimes strictly intervene with the client and insist on something or keep pressing on a point until a client is willing to look at it [2]. The process of human coaching includes an insight into how people learn and think, along with an understanding of what motivates them to achieve continuous high performance during behavioral intervention. Several coaching methods for the promotion of behavioral intervention are described in Table 5 [1-5,7,9,21-23,57,59,67,69,70,73]. The answer to "RQ3" contributes to "RQ4" and "RQ5" to analyze what limitations to overcome and what methods of offline behavioral intervention to include in eCoaching for the promotion of a healthy lifestyle.

Table 5. Coaching methods.

Method	Description
Systematic observation [7,23]	Systematic observation helps researchers to identify the instructional behaviors utilized by coaching practitioners within the practice environment. Observable and measurable data have the potential to solidify the scientific basis of the coaching process. Systematic observation must be capable of accurately and comprehensibly recording human behavior within a human coaching context.
Interpretive interview [2,3]	Achievement of the coaching process remains with observational data collection supplied with in-depth interviews that allow for the acquisition and interpretation of rich qualitative data based on the behavioral strategies of coaching.
Knowledge exchange [5]	In the search for an understanding of the coaching process, it is necessary to analyze and investigate the shared experience between the coach and participant.
Pragmatism [3]	Coaching is not a collection of techniques to apply or dogma to adhere to, rather it is a discipline that requires freshness, innovation, and relentless correction according to the outcomes being produced.
Understanding of human psychology [7,69,73]	Psychological principles on which coaching is based are essential. Without psychological understanding, coaches might go through the motions of coaching or use the behaviors associated with coaching, such as questioning, but fail to achieve the intended results.
Experience [5,7,67]	Experience is a skill that helps to improve competence and coaching outcomes, such as future advancement.
Trust [1]	Trust is one of the complex issues for coaches, whether internal and external. It teaches how not to use personal information and not to disclose it to illegitimate people.
Relationship [1]	The relationship must be based on mutual respect, trust, and mutual freedom of speech.
Expression [1]	Language impacts the goals of coaching by providing a means to assist the participant in being self-correcting and self-generating. It is important to provide new language to the participant for better understanding and learning.
Mentoring [3,5]	Mentoring is a more formal process, based on a one-to-one relationship with someone in the organization. While a mentor can use all the coaching types, their purpose is broader in scope than that of a coach.
Values and motivation [1]	Values are ideas about what is good and bad, and how things should be. Motivation is the internally generated feeling that stimulates participants to act. Motivation is related to the needs and values that have a correlation with intrinsic motivation.
Feedback [1,4]	Feedback is important for coaches to improve their learning environment.
Evidence based [4,7]	Evidence-based life coaching can enhance health, quality of life, and goal achievement.
Contextual [4,7,69]	Understanding the context is essential in coaching perspective, as it gives insights into why many participants either fail to use or resist the coaching approach.
Decision making [2,5,7]	Decision making includes data collection related to coaching, the privacy of the collected data, data cleaning, statistical analysis on the collected data, and the development of a machine learning model for prediction or regression analysis.
Goal based (goal setting) and evaluation [3,5,23,59]	Goals must be stated and measurable. Goals include clearly stated pathways to the preferred alternative by identifying strategies. Goal setting and goal evaluation are two essential parts of a behavioral intervention to determine the effectiveness of coaching. Goals must be specific, measurable, actionable, relevant, and time related. Evaluation of goals is important to understand the strengths and limitations of participants to set further attainable goals when necessary and reach the objectives.
Self-efficacy [9]	Self-efficacy has its core in social learning theory. It can be explained as the general or definite belief that people have concerning their capability to accomplish assigned tasks.
Personalization [21]	The concept of personalization or user tailoring is used in coaching to explain the variation in preferences between groups of participants and within the groups of participants to make recommendations more effective.
Persuasion [57]	Persuasion is a process that has been designed to change negative attitudes or behaviors of participants through advice, faith, and social influence. It is regularly used in the domain of public health where human-human or human-computer interaction is applied. It can be categorized as instruction style (authoritative and nonauthoritative), social feedback (cooperative and competitive), motivation type (extrinsic and intrinsic), and reinforcement type (negative and positive) [22].
Interaction and co-creation [70]	People are subject to self-regulation failures as follows: cravings, distractions, and deferring the right things. Therefore, people may need guidance through an eCoaching process to achieve the intended goal. Interaction is an integral part of pervasive computing that guides people to “do the right thing.” It requires improving automated logging of health (behavior) data and integrating this into coaching processes, as well as designing more intelligent and interactive coaching processes that incorporate user preferences and plans, contextual/situational priorities, and health data consequences. For successful design, the concept of co-creation or co-design is essential, where the system is designed together with its users.

RQ4: How Can the Methods of Human Coaching Processes Be Incorporated Into eCoaching for

Behavioral Intervention to Promote a Healthy Lifestyle?

The concept of eCoaching is constructed on the foundation of traditional coaching, and the technological revolution has boosted its performance and real-time acceptance. The World Health Organization (WHO) [49] claimed that chronic illnesses associated with modifiable lifestyle factors would be responsible for premature death worldwide. Therefore, change in negative health behavior should be given priority to avoid considerable losses caused by lifestyle diseases. An eCoach system can empower human beings to manipulate a healthy lifestyle with early health risk prediction and beneficial customized recommendations [22,61]. The pillars of eCoaching for behavioral intervention [56] are mostly inspired by the coaching methods as described in Table 5. They consist of data collection, data storage, analysis of data, goal setting, recommendation generation (intervention), monitoring, data privacy and ethics, goal evaluation, credibility, co-creation, feedback generation, and model evaluation [24,27,32,34,35,74]. Behavioral intervention is the process of intervening. As defined by WHO [49], a health intervention is an act performed for, with, or on behalf of a person or population, whose purpose is to assess, improve, maintain, promote, or modify health, functioning, or health conditions. Health interventions are used to promote a healthy lifestyle. Lifestyle or behavioral interventions include exercise, diet, and at least one other method (counseling, stress management, or healthy habits). Effective intervention planning is essential for an eCoach system for behavioral intervention to promote a healthy lifestyle change.

From the included "eCoach" articles, we found that the following methods are most appropriate for eCoaching processes: "personalization" (n=19) [21] "interaction and co-creation" (n=17) [70], "behavior change with technology" (n=17) [58], "goal setting" [59] and "evaluation" (n=16) [23,59], "persuasion" (n=15) [57], "automation" (n=14) [1], and "promotion of a healthy lifestyle" (n=14) [21]. These are relevant methods for eCoaching following a top-down ranking. "Personalized" recommendations are required to make intervention plans effective, and for that, personal "interaction" is necessary. For efficacy evaluation of eCoaching, personalized goal setting and goal evaluations are important. "Automation" is relevant to deliver automatic behavioral recommendations ("persuasion") to participants for the promotion of a "healthy lifestyle."

Methods in eCoaching, such as personalization, persuasion, goal setting and evaluation, interaction, and co-creation, are borrowed from traditional offline human coaching (Table 5). In eCoaching, persuasion is developed by trusting self-report or automation that observes human behavior using sensors, which is followed by health risk prediction with pattern recognition algorithms. The remaining four are core eCoaching methods. The first aspect is *automation* [1]. It helps to deliver automatic behavioral recommendations to users to maintain a healthy lifestyle. The decision support system (DSS) [22,61] within an eCoach system periodically monitors health and wellness parameters collected over time through sensors, questionnaires, and feedback forms, and predicts health risks. Once risk prediction is made, the DSS sends an automatic alert or recommendation to users. The second aspect is *behavior change with technology* [58]. Recent advancements in ICT have improved personal health care. The health care segment is still looking for an interactive, easy-to-use, optimized, cost-effective, and secure eCoach system for behavioral intervention for the promotion of a healthy lifestyle. The system should have the capability to normalize different formats of personalized data with appropriate ontological studies, ensuring the privacy of data. It should use AI algorithms based on ethical principles to analyze human psychology, monitor human behavior, and guide participants accordingly. Technology can support an eCoach by supporting coaching types, process management, human-computer interaction, remote collaborative work and communication, data collection and storage, data security and privacy, data analysis, recommendation generation, evaluation, and self-tracking. The third aspect is *promotion of a healthy lifestyle* [21]. Good health is the result of a healthy lifestyle, where caring about physical activities and nutrition are vital concerns. However, today, nutritional disorders are increasing rapidly. It is affecting children, adults, and older people, mainly due to limited nutrition knowledge and the lack of a healthy lifestyle. A commonly adopted approach for these imbalances is monitoring physical activity and daily habits, such as recording exercise and creating custom meal plans to count the number of macronutrients and micronutrients acquired in each meal. Behavioral interventions (nutritional and physical exercise coaching) through eCoaching have become popular (eg, Food4Living [17], TrainME [17], and RunningCoach [21]) for the promotion of a healthy lifestyle.

RQ5: How Can eCoaching Promote a Healthy Lifestyle With Proven Coaching Methods Using ICT?

The point of interest of eCoach initiatives is to deliver high-quality, evidence-based, comfortable, cost-effective, and timely care to assist human beings in retaining a healthy way of life [1,23,57]. eCoaching methods represent an evolving trend as they diverge from the conventional methods that tend to devalue user behavior. Health eCoaching is a complex process that demands careful planning and cooperation of several scientific domains, such as psychology, computer science, ethics, and medical science [23]. An effective eCoach design focuses on co-creation, co-design, and personalization of the intervention by the user and the system [23,30,32,35,58]. There are six primary attributes when modeling an eCoach system as follows [24,27,32,34,35,58,59,74]: (1) identification of the target group of participants, (2) selection of the study case, (3) type of data to be collected and data collection method, (4) target of coaching, (5) approach of coaching, and (6) evaluation of the intervention plan.

According to the findings in studies on coaching regarding the importance of the personal relationship between a coach and trainee, personalization of coaching strategies, motivation, goal setting, and engagement of the eCoach with the trainee/coached citizen/patient has to be customizable and easily available. Therefore, the user interface design of an eCoach system must be unambiguous and easily understandable [23,24], and it must not include unwanted artifacts. It must be designed following a standard co-creation process. eCoaching systems are an emerging trend with a design criterion to reduce the involvement

of human specialists with AI-inspired algorithms and robots for decision making based on supervised, unsupervised, and reinforced learning. In contrast, in several eCoach designs for behavioral intervention, human therapists or doctors, or other coaching experts are included [23,27,34,62]. The experts have access to the observation data, and they get involved or contribute to the coaching process.

eCoaching has other possibilities when compared to traditional coaching in terms of value addition, performance, and competence. Efficacy [52] study is a problem in both kinds of coaching to date, as revealed in the systematic literature review, for the following reasons: insufficient planning in study selection and study design [1,23], lack of conceptual/contextual clarity [24,25], inappropriate selection of sample size for statistical analysis [61], dearth of proper background education [74], lack of reliance and self-disclosure [1,22], absence of variation in a selected population [22,61], and lack of competence and experience with technology (digital illiteracy) [22,75].

Discussion

Overview

From the systematic literature review, we analyzed existing well-established traditional human coaching processes, descriptive models and the application domain, psychological approaches to describe coaching process models, methods in the coaching processes, and their applicability in eCoaching with the advancement of ICT to promote a healthy lifestyle. In this section, we discuss the findings associated with each individual research question.

Discussion on RQ1

The answer to "RQ1" helped us to identify key methods in the coaching process, as defined in Table 6 [1-7,9,16,66-68]. We studied their significance in eCoaching. The identified key methods in the coaching process are based on a review of established coaching process description models relevant for this RQ. Identified coaching methods are used to answer "RQ4." Appropriate coaching skills, knowledge to coach, method selection, proper implementation, personal interaction, and idea exchange are part of an effective coaching practice.

Table 6. Key methods associated with coaching.

Research group	Key methods associated with coaching
Potrac et al [66]	Behaviors, actions, and motivations
Cunningham et al [9]	Experience and racial diversity
Bartlett [1]	Trust, language, practice, and behavior
Côté [5]	Coach education and learning
Green et al [7]	Goal, psychology, evidence based, and cognition
Murphy [6]	Mentoring, evaluation, and leadership
Richards [67]	Intelligent coaching, learning, innovation, adaptation, Sustainability, and model performance
Flaherty [2]	Constraints of learning
Cox [3]	Pragmatism, experiencing, listening, clarifying, reflecting, and questioning
Stober et al [4]	Psychology, contextual, goal focus, cognition, and humanistic perspective
Knight [68]	Instructional coaching and visible learning
Standing et al [16]	Coaching, data collection, statistical analysis, and performance evaluation

Discussion on RQ2

We depicted the top three coaching descriptive models from Table 3, such as the models of Cavanagh et al [71], Whitten [8], and Cox [3] on "dynamic coaching," "cognitive-behavioral coaching," and "pragmatic inquiry into the coaching process," respectively. These models seemed to be suitable candidates to construct a personalized eCoaching process model for behavioral intervention for the promotion of a healthy lifestyle, contributing to the answer of "RQ4."

Discussion on RQ3

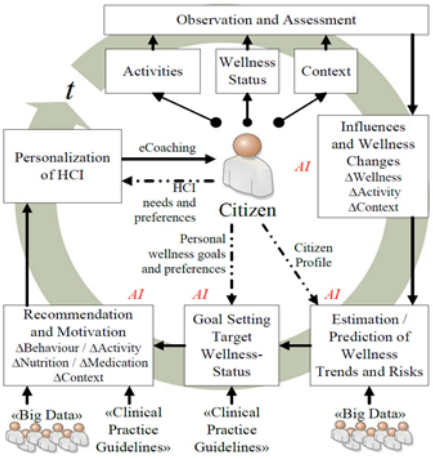
Systematic observation methods are recognized as useful research tools for providing quantitative descriptions of coaching behavior. Furthermore, coaching psychology mechanisms are also relevant for enhancing well-being, work performance, and personal life [7]. Therefore, researchers need to use systematic observation [64] and psychological coaching [7] to study coaching behavior in order to establish a database of meaningful coaching behaviors in different contexts. Besides the discussed strengths and potentials, constraints related to coaching as behavioral interventions are reviewed in the following text. *Language* is a medium of communication between people. Coaching may lead to language interpretability issues when selecting inappropriate language. Without proper communication, participants will be unable to perform the needed or desired tasks [2,66]. Regarding *understanding*, the humanistic nature of the coaching process remains little understood and an underresearched area [66]. Regarding *ethical dilemma*, researchers need to develop an ethical standard, adequate training, and presential coaching within a specific context. In many cases, coaches have not fully understood the

performance-related psychological principles on which coaching is based [69]. Regarding *diversity*, the coaching study should consider diversity in all its forms, such as organizational and occupational tenure, age, race, educational background, attitude, and personality [9]. Regrading *human behavior*, many coaches do not ground their practice in behavioral science. Participants should be selected from a diverse community, as members of a single community cannot represent the general population [7]. Regarding *conceptual clarity*, besides the popularity, the human coaching process reveals lack of conceptual clarity, imprecise description, and paucity of efficacy studies [6]. There persists a wide gap between what practitioners believe coaching is and what many executives think about coaching [67]. Regarding *implementation challenge*, the most formidable challenge in the field of coaching is the challenge of translating research into practice. Thought needs to be given to the sharing of all visible learning aspects in a way that is manageable and a part of goal-directed learning [68]. Regarding *bias*, due to background and bias, experts do ignore psychological problems they do not understand and may worsen the intervention. Thus, psychotherapeutic intervention is essential [4]. Regarding *human psychology and pressure*, most coaching-related studies are inclined to psychology rather than the way to do coaching. Pressure-based coaching hampers team functioning by negatively influencing team loyalty through increased levels of tension within the group [48].

Discussion on RQ4

The answer to "RQ2" revealed that the coaching model could be implemented in the following two ways: (1) the coach at the center and participants ("citizens") around, and (2) participants at the center and the coach around. Gerdes et al [61] proposed an eCoach concept based on monitoring, quantification of data, and AI, emphasizing human-centered design, with participants placed at the center, as depicted in Figure 5 [61]. The loop of the pictured eCoach model can be closed with an effective behavioral intervention plan, based on the selection of study cases, to guide people and deliver contextual and personalized recommendations to maintain a healthy lifestyle. This systematic literature review can help us understand how to solve the "What" (to coach) and "How" (to coach) questions related to eCoaching. In the eCoach model, as illustrated in Figure 5, the critical methods of coaching, as depicted in Figures 2-4, fit together for behavioral intervention.

Figure 5. A holistic electronic coach (eCoach) model proposed by Gerdes et al. AI: artificial intelligence; HCI: human-computer interaction.



Discussion on RQ5

Digital methods [17] for behavioral intervention with personal coaching have emerged as effective and scalable options. They include methods for intensive behavioral counseling, supporting face-to-face consultations with accessibility, attractive and personalized interaction, efficient use of time, and managing costs. eCoaching has the potential to overcome problems, such as language, bias, conceptual clarity, ill-defined matters, freedom of expression, pressure, and tension, which are expected in traditional coaching, as discussed in the answer to "RQ3." A smart eCoach may ideally deliver solutions asynchronously and on-demand with better flexibility and increased accessibility for personalized context-based coaching services. In this review, we have identified the following critical elements for effective eCoaching with AI [22,57,61,75-77]: real-time feedback, suggestion, and alert generation; preference sharing; comprehensible user experience (UX) design; interactive interaction (eg, intelligent chatbot); DSS; wellness vision (physical/social/emotional/spiritual); encouragement based on positive human psychology; assessment of human behavior based on physiological and contextual data; credibility; ethics; digital literacy to make the human-eCoach interaction effective; and generation of automatic, personalized, and context-specific recommendations to achieve health and wellness goals.

Strengths and Limitations

This systematic literature review helps to identify key processes from coaching methods to solve existing challenges and use human coaching and eCoaching processes as behavioral interventions. Coaching consists of observation, offering hints, feedback, reminders, and new tasks and redirecting participant attention to a salient goal to enhance participant performance. Coaching is applied to unveil the potential of participants to maximize their performance. A coach facilitates experimental learning that results in future-oriented abilities. A coach can shape new visions and plans to generate desired results. Despite the underutilization of remote coaching of health and wellness in the health care system as an asset of clinical practice, the focus on coaching is justified. The integration of offline human

coaching methods into the eCoaching process faces challenges related to privacy, ethics, coaching environment, skills, trust, motivation, intervention plans to change negative behaviors, human centeredness, and evaluation of preset goals. Despite the challenges, it is very promising to integrate human coaching methods into the eCoaching process [4,5,60]. An important limitation of this study is that we did not search the JMIR database, which has e-collections on the present topic. Future studies on this topic should search the JMIR database. This study serves as a basis for further research with a focus on designing an eCoach system based on the identified key coaching methods for the generation of personalized recommendations to achieve personal wellness goals.

Conclusion

An ideal coach should have the potential to conceptualize and navigate through changing complex environments. The coaching process is adopted to bridge inadequacies in areas where human resource structures and practices should play a more active mediating role. The success of the coaching process is an art, and impact analysis is important to evaluate its accomplishment. An evaluation of the human coaching process is also necessary. Therefore, the learning environment of active coaches needs to be continuously revisited and adapted. Health monitoring and fitness coaching with AI has the potential to contribute to research in eHealth. An optimized system for health eCoaching and management of personal health data that ensures data protection and privacy are significant challenges associated with eCoach-related research. The prediction of human behavior by analyzing human psychology for the generation of useful lifestyle recommendations is another challenging task to overcome, as human behavior is continuously changing. This review will provide eHealth researchers with an overview of different coaching and eCoaching processes, with the aim to promote a healthy lifestyle. In addition, this review can be used as a basis for further research focusing on the design, development, testing, and evaluation of the performance of an eCoach in order to generate automatic, meaningful, evidence-based, contextual, and personalized recommendations to achieve personal health goals.

Acknowledgments

We acknowledge funding and infrastructure from the University of Agder, Norway, to carry out this research. We thank the authors of previous studies for giving us consent to depict their coaching models in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search results from electronic databases.
[XLSX File (Microsoft Excel File), 12 KB - jmir_v23i3e23533_app1.xlsx]

References

1. Bartlett JE. Advances in coaching practices: A humanistic approach to coach and client roles. Journal of Business Research 2007 Jan;60(1):91-93. [doi: 10.1016/j.jbusres.2006.09.011]
2. Flaherty J. Coaching: Evoking Excellence in Others (2nd edition). Dev and Learning in Org 2006 Nov;20(6). [doi: 10.1108/dlo.2006.08120fae.002]
3. Cox E. Coaching Understood: A Pragmatic Inquiry into the Coaching Process. Thousand Oaks, CA, USA: SAGE Publications Ltd; 2013.
4. Stober D, Grant A. Evidence Based Coaching Handbook: Putting Best Practices to Work for Your Clients. Hoboken, NJ, USA: John Wiley & Sons; 2010.
5. Côté J. The Development of Coaching Knowledge. International Journal of Sports Science & Coaching 2016 Nov 07;1(3):217-222. [doi: 10.1260/174795406778604609]
6. Murphy SA. Recourse to Executive Coaching: The Mediating Role of Human Resources. International Journal of Police Science & Management 2016 Jul 31;7(3):175-186. [doi: 10.1350/ijps.2005.7.3.175]
7. Green LS, Oades LG, Grant AM. Cognitive-behavioral, solution-focused life coaching: Enhancing goal striving, well-being, and hope. The Journal of Positive Psychology 2006 Jul;1(3):142-149. [doi: 10.1080/17439760600619849]
8. Whitten H. Cognitive Behavioural Coaching Techniques For Dummies. Hoboken, NJ, USA: John Wiley & Sons; 2011.
9. Cunningham GB, Sagas M. People make the difference: The influence of the coaching staff's human capital and diversity on team performance. European Sport Management Quarterly 2004 Mar;4(1):3-21. [doi: 10.1080/16184740408737464]
10. Grant A, Palmer S. Integrating positive psychology and coaching psychology into counselling psychology. Counselling Psychology Review 2015 Sep;30(3):22-25 [FREE Full text]
11. Rutten GM, Meis JJM, Hendriks MRC, Hamers FJM, Veenhof C, Kremers SPJ. The contribution of lifestyle coaching of overweight patients in primary care to more autonomous motivation for physical activity and healthy dietary behaviour: results of a longitudinal study. Int J Behav Nutr Phys Act 2014 Jul 16;11:86 [FREE Full text] [doi: 10.1186/s12966-014-0086-z] [Medline: 25027848]

12. Brötz D. Exercise, Physical Activity, and Self-Determination Theory: A Systematic Review. physioscience 2013 Feb 26;9(01):37-37 [FREE Full text] [doi: 10.1055/s-0032-1330638]
13. Venditti EM, Wylie-Rosett J, Delahanty LM, Mele L, Hoskin MA, Edelstein SL, Diabetes Prevention Program Research Group. Short and long-term lifestyle coaching approaches used to address diverse participant barriers to weight loss and physical activity adherence. Int J Behav Nutr Phys Act 2014 Feb 12;11:16 [FREE Full text] [doi: 10.1186/1479-5868-11-16] [Medline: 24521153]
14. Gillison F, Stathi A, Reddy P, Perry R, Taylor G, Bennett P, et al. Processes of behavior change and weight loss in a theory-based weight loss intervention program: a test of the process model for lifestyle behavior change. Int J Behav Nutr Phys Act 2015 Jan 16;12:2 [FREE Full text] [doi: 10.1186/s12966-014-0160-6] [Medline: 25592314]
15. Lakerveld J, Bot SD, Chinapaw MJ, van Tulder MW, Kostense PJ, Dekker JM, et al. Motivational interviewing and problem solving treatment to reduce type 2 diabetes and cardiovascular disease risk in real life: a randomized controlled trial. Int J Behav Nutr Phys Act 2013 Apr 19;10:47 [FREE Full text] [doi: 10.1186/1479-5868-10-47] [Medline: 23597082]
16. Standing R, Maulder P. The Effectiveness of Progressive and Traditional Coaching Strategies to Improve Sprint and Jump Performance Across Varying Levels of Maturation within a General Youth Population. Sports (Basel) 2019 Jul 30;7(8):186 [FREE Full text] [doi: 10.3390/sports7080186] [Medline: 31366104]
17. Castro Sweet CM, Chiguluri V, Gumpina R, Abbott P, Madero EN, Payne M, et al. Outcomes of a Digital Health Program With Human Coaching for Diabetes Risk Reduction in a Medicare Population. J Aging Health 2018 Jun 01;30(5):692-710 [FREE Full text] [doi: 10.1177/0898264316688791] [Medline: 28553807]
18. den Braber N, Vollenbroek-Hutten MMR, Oosterwijk MM, Gant CM, Hagedoorn IJM, van Beijnum BF, et al. Requirements of an Application to Monitor Diet, Physical Activity and Glucose Values in Patients with Type 2 Diabetes: The Diameter. Nutrients 2019 Feb 15;11(2):409 [FREE Full text] [doi: 10.3390/nu11020409] [Medline: 30781348]
19. Digital health. World Health Organization. URL: https://www.who.int/ehealth/en/ [accessed 2020-07-17]
20. Chatterjee A, Gerdes MW, Martinez SG. Identification of Risk Factors Associated with Obesity and Overweight-A Machine Learning Overview. Sensors (Basel) 2020 May 11;20(9):2734 [FREE Full text] [doi: 10.3390/s20092734] [Medline: 32403349]
21. Aranki D, Peh GX, Kurillo G, Bajcsy R. The Feasibility and Usability of RunningCoach: A Remote Coaching System for Long-Distance Runners. Sensors (Basel) 2018 Jan 10;18(1):175 [FREE Full text] [doi: 10.3390/s18010175] [Medline: 29320436]
22. Chatterjee A, Gerdes M, Martinez S. eHealth Initiatives for The Promotion of Healthy Lifestyle and Allied Implementation Difficulties. 2019 Oct Presented at: 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob); October 21-23, 2019; Barcelona, Spain. [doi: 10.1109/wimob.2019.8923324]
23. Petsani D, Kostantinidis E, Bamidis P. Designing an E-coaching System for Older People to Increase Adherence to Exergame-based Physical Activity. In: Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health. 2018 Presented at: 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health; 2018; Portugal p. 258-263. [doi: 10.5220/0006821502580263]
24. Benítez-Guijarro A, Ruiz-Zafra Á, Callejas Z, Medina-Medina N, Benghazi K, Noguera M. General Architecture for Development of Virtual Coaches for Healthy Habits Monitoring and Encouragement. Sensors (Basel) 2018 Dec 30;19(1):108 [FREE Full text] [doi: 10.3390/s19010108] [Medline: 30598011]
25. Montenegro C, López Zorrilla A, Mikel Olaso J, Santana R, Justo R, Lozano JA, et al. A Dialogue-Act Taxonomy for a Virtual Coach Designed to Improve the Life of Elderly. MTI 2019 Jul 11;3(3):52. [doi: 10.3390/mti3030052]
26. Hansel B, Giral P, Gambotti L, Lafourcade A, Peres G, Filipecki C, et al. A Fully Automated Web-Based Program Improves Lifestyle Habits and HbA1c in Patients With Type 2 Diabetes and Abdominal Obesity: Randomized Trial of Patient E-Coaching Nutritional Support (The ANODE Study). J Med Internet Res 2017 Nov 08;19(11):e360 [FREE Full text] [doi: 10.2196/jmir.7947] [Medline: 29117929]
27. Beun RJ, Fitrianie S, Griffioen-Both F, Spruit S, Horsch C, Lancee J, et al. Talk and Tools: the best of both worlds in mobile user interfaces for E-coaching. Pers Ubiquit Comput 2017 May 19;21(4):661-674. [doi: 10.1007/s00779-017-1021-5]
28. Boratto L, Carta S, Mulas F, Pilloni P. An e-coaching ecosystem: design and effectiveness analysis of the engagement of remote coaching on athletes. Pers Ubiquit Comput 2017 May 23;21(4):689-704. [doi: 10.1007/s00779-017-1026-0]
29. Kamphorst B, Kalis A. Why option generation matters for the design of autonomous e-coaching systems. AI & Soc 2014 Jan 17;30(1):77-88. [doi: 10.1007/s00146-013-0532-5]
30. Anaby D, Mercerat C, Tremblay S. Enhancing Youth Participation Using the PREP Intervention: Parents' Perspectives. Int J Environ Res Public Health 2017 Sep 02;14(9):1005 [FREE Full text] [doi: 10.3390/ijerph14091005] [Medline: 28869491]
31. Lim S, Oh H, Kim J, Lee J, Park J. LSTM-Guided Coaching Assistant for Table Tennis Practice. Sensors (Basel) 2018 Nov 23;18(12):4112 [FREE Full text] [doi: 10.3390/s18124112] [Medline: 30477175]
32. Klaassen R, Bul KCM, Op den Akker R, van der Burg GJ, Kato PM, Di Bitonto P. Design and Evaluation of a Pervasive Coaching and Gamification Platform for Young Diabetes Patients. Sensors (Basel) 2018 Jan 30;18(2):402 [FREE Full text] [doi: 10.3390/s18020402] [Medline: 29385750]
33. Nooitgedagt A, Beun R, Dignum F. e-Coaching for Intensive Cardiac Rehabilitation. In: de Vries P, Oinas-Kukkonen H, Siemons L, Beerlage-de Jong N, van Gemert-Pijnen L, editors. Persuasive Technology: Development and Implementation

of Personalized Technologies to Change Attitudes and Behaviors. PERSUASIVE 2017. Lecture Notes in Computer Science, vol 10171. Cham: Springer; 2017:31-42.
34. Beun R, Brinkman W, Fitrianie S. Improving Adherence in Automated e-Coaching. In: Meschtscherjakov A, De Ruyter B, Fuchsberger V, Murer M, Tscheligi M, editors. Persuasive Technology. PERSUASIVE 2016. Lecture Notes in Computer Science, vol 9638. Cham: Springer; 2016:276-287.
35. Pratiwi P, Tjondronegoro D. Towards personalisation of physical activity e-coach using stage-matched behaviour change and motivational interviewing strategies. 2017 Presented at: 2017 IEEE Life Sciences Conference (LSC); December 13-15, 2017; Sydney, Australia p. 5-8. [doi: 10.1109/lsc.2017.8268130]
36. Nápoles AM, Santoyo-Olsson J, Chacón L, Stewart AL, Dixit N, Ortiz C. Feasibility of a Mobile Phone App and Telephone Coaching Survivorship Care Planning Program Among Spanish-Speaking Breast Cancer Survivors. JMIR Cancer 2019 Jul 09;5(2):e13543 [FREE Full text] [doi: 10.2196/13543] [Medline: 31290395]
37. Partridge SR, Allman-Farinelli M, McGeechan K, Balestracci K, Wong ATY, Hebden L, et al. Process evaluation of TXT2BFiT: a multi-component mHealth randomised controlled trial to prevent weight gain in young adults. Int J Behav Nutr Phys Act 2016 Jan 19;13(1):7 [FREE Full text] [doi: 10.1186/s12966-016-0329-2] [Medline: 26785637]
38. Middelweerd A, van der Laan DM, van Stralen MM, Mollee JS, Stuij M, te Velde SJ, et al. What features do Dutch university students prefer in a smartphone application for promotion of physical activity? A qualitative approach. Int J Behav Nutr Phys Act 2015 Mar 01;12:31 [FREE Full text] [doi: 10.1186/s12966-015-0189-1] [Medline: 25889577]
39. Partridge SR, McGeechan K, Bauman A, Phongsavan P, Allman-Farinelli M. Improved eating behaviours mediate weight gain prevention of young adults: moderation and mediation results of a randomised controlled trial of TXT2BFiT, mHealth program. Int J Behav Nutr Phys Act 2016 Apr 02;13:44 [FREE Full text] [doi: 10.1186/s12966-016-0368-8] [Medline: 27039178]
40. Müller AM, Khoo S. Non-face-to-face physical activity interventions in older adults: a systematic review. Int J Behav Nutr Phys Act 2014 Mar 10;11(1):35 [FREE Full text] [doi: 10.1186/1479-5868-11-35] [Medline: 24612748]
41. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
42. Greaves C, Gillison F, Stathi A, Bennett P, Reddy P, Dunbar J, et al. Waste the waist: a pilot randomised controlled trial of a primary care based intervention to support lifestyle change in people with high cardiovascular risk. Int J Behav Nutr Phys Act 2015 Jan 16;12:1 [FREE Full text] [doi: 10.1186/s12966-014-0159-z] [Medline: 25592201]
43. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug 11;14(1):105 [FREE Full text] [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]
44. Michaelides A, Major J, Pienkosz E, Wood M, Kim Y, Toro-Ramos T. Usefulness of a Novel Mobile Diabetes Prevention Program Delivery Platform With Human Coaching: 65-Week Observational Follow-Up. JMIR Mhealth Uhealth 2018 May 03;6(5):e93 [FREE Full text] [doi: 10.2196/mhealth.9161] [Medline: 29724709]
45. Brandt CJ, Søgaard GI, Clemensen J, Søndergaard J, Nielsen JB. Determinants of Successful eHealth Coaching for Consumer Lifestyle Changes: Qualitative Interview Study Among Health Care Professionals. J Med Internet Res 2018 Jul 05;20(7):e237 [FREE Full text] [doi: 10.2196/jmir.9791] [Medline: 29980496]
46. Mao AY, Chen C, Magana C, Caballero Barajas K, Olayiwola JN. A Mobile Phone-Based Health Coaching Intervention for Weight Loss and Blood Pressure Reduction in a National Payer Population: A Retrospective Study. JMIR Mhealth Uhealth 2017 Jun 08;5(6):e80 [FREE Full text] [doi: 10.2196/mhealth.7591] [Medline: 28596147]
47. Karhula T, Vuorinen A, Rääpysjärvi K, Pakanen M, Itkonen P, Tepponen M, et al. Telemonitoring and Mobile Phone-Based Health Coaching Among Finnish Diabetic and Heart Disease Patients: Randomized Controlled Trial. J Med Internet Res 2015 Jun 17;17(6):e153 [FREE Full text] [doi: 10.2196/jmir.4059] [Medline: 26084979]
48. Wayne N, Perez DF, Kaplan DM, Ritvo P. Health Coaching Reduces HbA1c in Type 2 Diabetic Patients From a Lower-Socioeconomic Status Community: A Randomized Controlled Trial. J Med Internet Res 2015 Oct 05;17(10):e224 [FREE Full text] [doi: 10.2196/jmir.4871] [Medline: 26441467]
49. Modave F, Bian J, Leavitt T, Bromwell J, Harris Iii C, Vincent H. Low Quality of Free Coaching Apps With Respect to the American College of Sports Medicine Guidelines: A Review of Current Mobile Apps. JMIR Mhealth Uhealth 2015 Jul 24;3(3):e77 [FREE Full text] [doi: 10.2196/mhealth.4669] [Medline: 26209109]
50. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. It's LiFe! Mobile and Web-Based Monitoring and Feedback Tool Embedded in Primary Care Increases Physical Activity: A Cluster Randomized Controlled Trial. J Med Internet Res 2015 Jul 24;17(7):e184 [FREE Full text] [doi: 10.2196/jmir.4579] [Medline: 26209025]
51. Verwey R, van der Weegen S, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. A monitoring and feedback tool embedded in a counselling protocol to increase physical activity of patients with COPD or type 2 diabetes in primary care: study protocol of a three-arm cluster randomised controlled trial. BMC Fam Pract 2014 May 12;15:93 [FREE Full text] [doi: 10.1186/1471-2296-15-93] [Medline: 24885096]

52. Karppinen P, Oinas-Kukkonen H, Alahäivälä T, Jokelainen T, Keränen AM, Salonurmi T, et al. Persuasive user experiences of a health Behavior Change Support System: A 12-month study for prevention of metabolic syndrome. Int J Med Inform 2016 Dec;96:51-61. [doi: 10.1016/j.ijmedinf.2016.02.005] [Medline: 26992482]
53. Karppinen P, Oinas-Kukkonen H, Alahäivälä T, Jokelainen T, Teeriniemi A, Salonurmi T, et al. Opportunities and challenges of behavior change support systems for enhancing habit formation: A qualitative study. J Biomed Inform 2018 Aug;84:82-92 [FREE Full text] [doi: 10.1016/j.jbi.2018.06.012] [Medline: 29936091]
54. Teeriniemi A, Salonurmi T, Jokelainen T, Vähänikkilä H, Alahäivälä T, Karppinen P, et al. A randomized clinical trial of the effectiveness of a Web-based health behaviour change support system and group lifestyle counselling on body weight loss in overweight and obese subjects: 2-year outcomes. J Intern Med 2018 Nov 04;284(5):534-545 [FREE Full text] [doi: 10.1111/joim.12802] [Medline: 29974563]
55. Kamphorst BA. E-coaching systems. Pers Ubiquit Comput 2017 May 23;21(4):625-632. [doi: 10.1007/s00779-017-1020-6]
56. Lentferink AJ, Oldenhuis HK, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JE. Key Components in eHealth Interventions Combining Self-Tracking and Persuasive eCoaching to Promote a Healthier Lifestyle: A Scoping Review. J Med Internet Res 2017 Aug 01;19(8):e277 [FREE Full text] [doi: 10.2196/jmir.7288] [Medline: 28765103]
57. Ochoa SF, Gutierrez FJ. Architecting E-Coaching Systems: A First Step for Dealing with Their Intrinsic Design Complexity. Computer 2018 Mar;51(3):16-23. [doi: 10.1109/mc.2018.1731079]
58. Fischer X, Donath L, Zwygart K, Gerber M, Faude O, Zahner L. Coaching and Prompting for Remote Physical Activity Promotion: Study Protocol of a Three-Arm Randomized Controlled Trial (Movingcall). Int J Environ Res Public Health 2019 Jan 25;16(3):331 [FREE Full text] [doi: 10.3390/ijerph16030331] [Medline: 30691013]
59. Pascal A, Sass M, Gregory JB. I'm only human: The role of technology in coaching. Consulting Psychology Journal: Practice and Research 2015;67(2):100-109. [doi: 10.1037/cpb0000025]
60. van Berkel J, Boot CRL, Proper KI, Bongers PM, van der Beek AJ. Effectiveness of a worksite mindfulness-related multi-component health promotion intervention on work engagement and mental health: results of a randomized controlled trial. PLoS One 2014;9(1):e84118 [FREE Full text] [doi: 10.1371/journal.pone.0084118] [Medline: 24489648]
61. Gerdes M, Martinez S, Tjondronegoro D. Conceptualization of a personalized ecoach for wellness promotion. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare - PervasiveHealth '17. 2017 Presented at: 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 2017; Barcelona, Spain p. 365-374. [doi: 10.1145/3154862.3154930]
62. Rutjes H, Willemsen MC, Kersten - van Dijk ET, de Ruyter B, IJsselsteijn WA. Better together: opportunities for technology in health coaching from the coach's perspective. 2017 Presented at: 2017 ECHW Conference; January 26-27, 2017; Amsterdam, Netherlands.
63. Kamphorst BA, Klein MCA, Van Wissen A. Autonomous e-coaching in the wild: empirical validation of a model-based reasoning system. In: AAMAS '14: Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems. 2014 Presented at: 2014 international conference on Autonomous agents and multi-agent systems; May 2014; Paris, France p. 725-732.
64. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
65. Wendler R. The maturity of maturity model research: A systematic mapping study. Information and Software Technology 2012 Dec;54(12):1317-1339. [doi: 10.1016/j.infsof.2012.07.007]
66. Potrac P, Brewer C, Jones R, Armour K, Hoff J. Toward an Holistic Understanding of the Coaching Process. Quest 2000 May;52(2):186-199. [doi: 10.1080/00336297.2000.10491709]
67. Richards B. Intelligent coaching: Unleashing human potential. The Journal for Quality and Participation 2003;26(1):13-16 [FREE Full text]
68. Knight J. Instructional Coaching for Implementing Visible Learning: A Model for Translating Research into Practice. Education Sciences 2019 May 10;9(2):101. [doi: 10.3390/educsci9020101]
69. Szymanska K. Anxiety and the coaching relationship: How to recognise the signs and what to do next. The Coaching Psychologist 2007;3(2):85-89 [FREE Full text]
70. Beun R, Griffioen-both F, Ahn R, Fitrianie S, Lancee J. Modeling interaction in automated e-coaching a case from insomnia therapy. In: COGNITIVE 2014: The Sixth International Conference on Advanced Cognitive Technologies and Applications. Wilmington, DE: IARIA; 2014.
71. Cavanagh M, Spence G. Mindfulness in Coaching. In: The Wiley ‐ Blackwell Handbook of the Psychology of Coaching and Mentoring. Hoboken, NJ, USA: Wiley; 2012:112-134.
72. Roša (Rosha) A, Lace N. The Open Innovation Model of Coaching Interaction in Organisations for Sustainable Performance within the Life Cycle. Sustainability 2018 Sep 30;10(10):3516. [doi: 10.3390/su10103516]
73. Guyenet SJ. The Hungry Brain: Outsmarting the Instincts That Make Us Overeat. New York, NY, USA: Macmillan; 2017.
74. Kovalchuck V, Vorotnykova I. E-Coaching, E-Mentoring for Lifelong Professional Development of Teachers within the System of Post-Graduate Pedagogical Education. Turkish Online Journal of Distance Education 2017:214. [doi: 10.17718/TOJDE.328956]

75. Shneiderman B, Plaisant C, Cohen M, Jacobs S, Elmqvist N, Diakopoulos N. Designing the User Interface: Strategies for Effective Human-Computer Interaction. London, United Kingdom: Pearson; 2016.
76. Baber C, Khattab A, Russell M, Hermsdörfer J, Wing A. Creating Affording Situations: Coaching through Animate Objects. Sensors (Basel) 2017 Oct 11;17(10):2308 [FREE Full text] [doi: 10.3390/s17102308] [Medline: 29019939]
77. Jablonski RA, Winstead V, Geldmacher DS. Description of Process and Content of Online Dementia Coaching for Family Caregivers of Persons with Dementia. Healthcare (Basel) 2019 Jan 19;7(1):13 [FREE Full text] [doi: 10.3390/healthcare7010013] [Medline: 30669444]

Abbreviations

AI: artificial intelligence
DSS: decision support system
eCoach: electronic coach
eCoaching: electronic coaching
ICT: information and communication technology
RQ: research question
WHO: World Health Organization

Edited by G Eysenbach, R Kukafka; submitted 14.08.20; peer-reviewed by N Nazari, M Vollenbroek-Hutten; comments to author 03.11.20; revised version received 28.12.20; accepted 15.02.21; published 24.03.21.

Please cite as:
Chatterjee A, Gerdes M, Prinz A, Martinez S
Human Coaching Methodologies for Automatic Electronic Coaching (eCoaching) as Behavioral Interventions With Information and Communication Technology: Systematic Review
J Med Internet Res 2021;23(3):e23533
URL: https://www.jmir.org/2021/3/e23533
doi:10.2196/23533
PMID:33759793

Review

Barriers to and Facilitators of User Engagement With Digital Mental Health Interventions: Systematic Review

Judith Borghouts[1], PhD; Elizabeth Eikey[2], PhD; Gloria Mark[1], PhD; Cinthia De Leon[1], MPH; Stephen M Schueller[1], PhD; Margaret Schneider[1], PhD; Nicole Stadnick[2], MPH, PhD; Kai Zheng[1], PhD; Dana Mukamel[1], PhD; Dara H Sorkin[1], PhD

[1]University of California Irvine, Irvine, CA, United States

[2]University of California San Diego, San Diego, CA, United States

Corresponding Author:
Judith Borghouts, PhD
University of California Irvine
6210 Donald Bren Hall
Irvine, CA, 92697-3425
United States
Phone: 1 9498240246
Email: jborghou@uci.edu

Abstract

Background: Digital mental health interventions (DMHIs), which deliver mental health support via technologies such as mobile apps, can increase access to mental health support, and many studies have demonstrated their effectiveness in improving symptoms. However, user engagement varies, with regard to a user's uptake and sustained interactions with these interventions.

Objective: This systematic review aims to identify common barriers and facilitators that influence user engagement with DMHIs.

Methods: A systematic search was conducted in the SCOPUS, PubMed, PsycINFO, Web of Science, and Cochrane Library databases. Empirical studies that report qualitative and/or quantitative data were included.

Results: A total of 208 articles met the inclusion criteria. The included articles used a variety of methodologies, including interviews, surveys, focus groups, workshops, field studies, and analysis of user reviews. Factors extracted for coding were related to the end user, the program or content offered by the intervention, and the technology and implementation environment. Common barriers included severe mental health issues that hampered engagement, technical issues, and a lack of personalization. Common facilitators were social connectedness facilitated by the intervention, increased insight into health, and a feeling of being in control of one's own health.

Conclusions: Although previous research suggests that DMHIs can be useful in supporting mental health, contextual factors are important determinants of whether users actually engage with these interventions. The factors identified in this review can provide guidance when evaluating DMHIs to help explain and understand user engagement and can inform the design and development of new digital interventions.

(J Med Internet Res 2021;23(3):e24387) doi:10.2196/24387

KEYWORDS

mHealth; eHealth; mental health; depression; anxiety; behavior; mobile phone

Introduction

Background

Nearly 1 in 5 adults in the United States experience a mental illness at some moment in their life [1]. Yet, accessing treatment for mental health problems can be difficult. Common barriers to mental health care include stigma, lack of available and evidence-based services, and inability to afford services [2,3]. In addition, people not diagnosed with a mental illness can experience periods of poor mental health and may benefit from support, although they have not sought professional treatment with a mental health provider. For instance, 73% of people surveyed in the United States experience stress related to money, work, and family responsibilities at a level that affects their mental health [4]. The translation of psychosocial interventions

into digital formats, deemed digital mental health interventions (DMHIs), has the potential to overcome some existing barriers to traditional care and increase access to mental health support and resources.

DMHIs can be delivered via smartphone apps, internet websites, wearable devices, virtual reality, or video games [5] and range from self-guided DMHIs to those integrated with human support or traditional therapy [6]. Although some DMHIs have been shown to be as effective as traditional mental health services (eg, psychotherapy and pharmacotherapy) in improving mental health conditions such as depression [7] and can lead to greater reductions in anxiety compared with usual care [8], engagement with these technologies remains to be an ongoing issue, varies from study to study, and is typically lower in real-world use than research studies [9]. For example, a review in 2018 found that participant adherence to internet-delivered cognitive behavioral therapy (CBT) can range from 6% to 100% [10]. Similarly, systematic comparisons in 2018 and 2019 on self-help DMHIs found that real-world uptake varies widely [9,11], and acceptability can be lower than traditional treatment [12].

This paper aims to systematically review the literature on DMHIs to identify common barriers and facilitators that may influence user engagement with these interventions. There are different ways to define user engagement. For example, engagement can be referred to as the time a user spends on an intervention. However, the time spent on an intervention varies between different types of interventions, and little time spent using a DMHI does not have to be a negative feature per se. To get a comprehensive understanding of people's use of DMHIs, we use a broader definition of user engagement. In this review, *user engagement* refers to a user's uptake and sustained interactions with a digital intervention, which includes interest in adopting an intervention as demonstrated by signing up for the digital intervention, initial uptake as demonstrated by engaging with features of the digital intervention as part of the study, at a minimum during a demonstration as part of the study, and continued use of an intervention.

Understanding User Engagement With DMHIs

A range of factors can influence engagement with DMHIs, such as the relevance of information to the user provided by a digital intervention [13], a lack of user motivation to persist with a self-guided intervention [14], and poor user experience with the technology [15]. Although previous studies have each reported on *some* factors that can influence engagement, given a particular technology or context, a review is lacking that brings all these findings together. It is important to investigate the multitude of factors to fully understand the reasons for high versus low engagement. Previous reviews have highlighted the variability in engagement and uptake, analyzing both DMHIs published in the academic literature [9-11] and publicly available mental health apps in app stores [16]. However, these analyses did not report on factors related to this variability in engagement. This review seeks to address this gap by identifying the most common overarching factors that affect engagement.

Although analyzing engagement metrics of commercial apps can be used to examine variability in engagement, user studies are valuable to understand the underlying reasons why people may engage with some interventions more than others. For the purpose of this review, we focus on reviewing the academic literature.

Researchers and developers of DMHIs can use this knowledge to inform evaluations of engagement and the development of new digital interventions. In addition, it may provide insights into what services and facilitating conditions need to surround DMHIs to promote technology-enabled services and may help mental health service providers in selecting suitable interventions for their clients.

This review focuses on common mental health issues, such as depression, anxiety, psychological well-being and distress, and stress. There may be different barriers or facilitators for user engagement with other specific, serious mental illness interventions (eg, psychosis intervention) that are beyond the scope of this paper.

Methods

Inclusion Criteria

The inclusion and exclusion criteria of articles for this review are presented in Textboxes 1 and 2, respectively.

Textbox 1. Inclusion criteria.

- Report on an intervention aimed to improve mental health, psychological well-being, anxiety, depression, stress, and/or mood
- Report on an intervention delivered in a digital format, such as a smartphone app or website
- Report on some aspects of user experience (eg, usability, user satisfaction, and user feedback)
- Report on factors that affected user experience
- Include participants aged ≥16 years (eg, child and adolescent samples were excluded)
- Report on an empirical study (eg, literature reviews that synthesized findings from other articles, columns, opinion pieces, comments or replies, and editorials were excluded)
- Be a peer-reviewed article (eg, dissertations were excluded)
- Be written in English

Textbox 2. Exclusion criteria.

- Report on interventions that have a mental health component but do not have mental health as a primary intervention target (eg, an app that is primarily focused on physical pain symptoms, with a mental health component)
- Report on interventions that only serve as an appointment booking system for in-person therapy
- Report on interventions that are used as a component during an in-person session but cannot be used remotely outside of these sessions
- Articles published before January 1, 2010

The first exclusion criterion was added to identify barriers and facilitators that would be applicable to DMHIs. For example, a study that tests an app primarily focused on physical pain symptoms, with a mental health component, may find physical pain issues as a barrier to engaging with the app. It may not be clear from the study whether this is a common barrier related to DMHIs or interventions addressing physical pain.

The second and third exclusion criteria were added, as these types of interventions were designed to be a part of in-person sessions. It may not be clear from these studies whether users would be willing or able to engage with DMHIs apart from existing and traditional in-person sessions.

Finally, digital health interventions evolve rapidly [17,18], and the review was focused on the current state of DMHIs. Therefore, to avoid discussing on interventions or technologies that are now potentially out of date, the review was limited to contemporary studies published within the last 10 years (January 2010 to December 2019), a time frame that has been applied previously for systematic reviews on digital health technologies for mental illness [18].

Search Strategy

A literature search was conducted in multiple databases, including SCOPUS, PubMed, PsycINFO, Web of Science, and the Cochrane Library. On the basis of the inclusion criteria, a search query was developed to include an article if its title or abstract contained at least one keyword related to mental health, at least one keyword related to digital interventions, and at least one keyword related to user experience (Textbox 3; PRE/5 means that keywords were separated by a maximum of 5 words, for example, online PRE/5 intervention means there were 5 or less words between *online* and *intervention*).

The search query was built on keywords used in previous reviews on the uptake of mental health technologies [11,19], and additional keywords were added for the specific focus of this review (ie, the third part of the query with keywords related to user experience). The search terms for each database are included in Multimedia Appendix 1. Searches were not limited to the study design.

Textbox 3. Search query.

TITLE-ABS-KEY (depress* OR anxiet* OR anxious OR mood OR “mental health”

OR “psychological wellbeing” OR “mental wellbeing” OR “behavioral health” OR

“mental illness”)

AND TITL E-ABS-KEY ((online PRE/5 intervention*) OR (online PRE/5

treatment) OR (digital PRE/5 intervention*) OR (digital PRE/5 treatment)

OR (mobile PRE/5 intervention*) OR (mobile PRE/5 treatment) OR (smartphone PRE/5 intervention*) OR (smartphone PRE/5 treatment) OR (web-based

PRE/5 intervention*) OR (web-based PRE/5 treatment) OR (internet PRE/5

intervention*) OR (internet PRE/5 treatment) OR (computer PRE/5 intervention*) OR (computer PRE/5 treatment) OR (cyber PRE/5 intervention*) OR (cyber PRE/5 treatment) OR (electronic PRE/5 intervention*) OR (electronic PRE/5 treatment) OR (mobile AND program*) OR mhealth OR ehealth OR mtherap* OR etherap* OR telehealth OR telemedicine OR “mobile app*”)

AND TITLE-ABS-KEY (usability OR “user experience” OR evaluation* OR

engagement OR interface OR satisfaction OR usage OR adoption OR acceptability OR qualitative OR user perspective* OR barrier* OR interview* OR focus group*)

Study Selection

The search results were uploaded to Rayyan [20], a web-based software program for facilitating systematic reviews. Titles and abstracts were screened against the inclusion criteria, and excluded articles were labeled with reasons for exclusion.

The first author reviewed all titles and abstracts. Explicit inclusion criteria were determined between the first 3 authors a priori article selection to reduce coder bias. The coder (JB) was a PhD researcher with years of research expertise in user experience and thematic analysis.

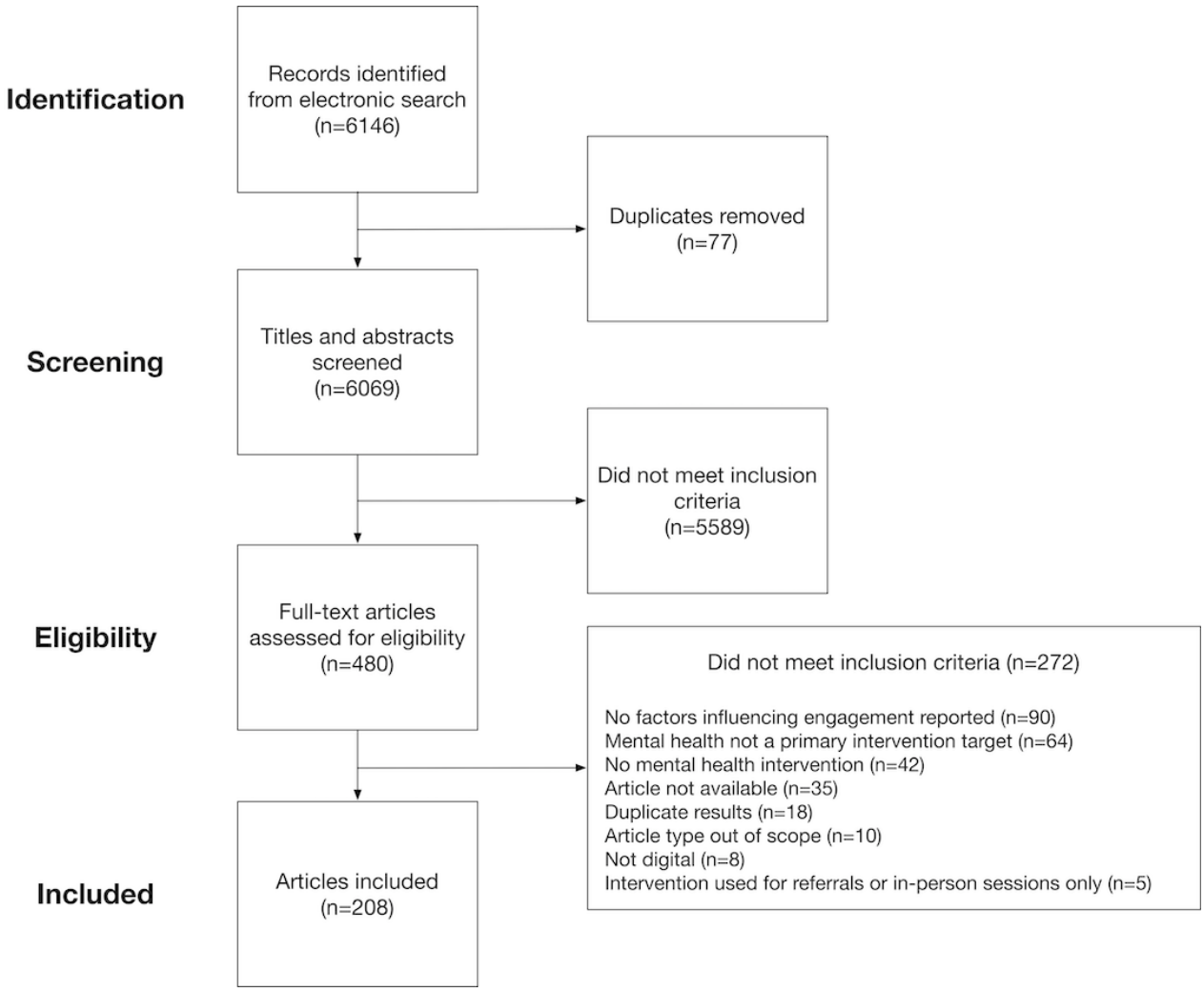
A total of 6146 papers were extracted for the review. After the removal of 77 duplicates, 6069 article titles and abstracts were screened by the first author and discussed with the second and third authors. Uncertainties about inclusion were resolved by discussion among the first 3 authors, and reasons for exclusion or inclusion of these articles were discussed.

Furthermore, 480 full-text articles were reviewed, of which 208 met the inclusion criteria. Figure 1 shows a flow diagram of the screening papers. The same inclusion and exclusion criteria were used for reviewing the titles and abstracts in the screening phase and reviewing the full-text articles in the eligibility phase.

Articles that were not available were either not available on the web or were behind a paid firewall. Article types that were out of scope did not report on an empirical study.

Although there is a risk of bias in studies, the review considered all studies that met the inclusion criteria and included a large variety of different study methodologies, including qualitative studies with no reported quantitative outcomes. The primary focus of this review was to establish themes across the literature rather than extract the outcomes of quantitative studies. Therefore, the risk of publication bias with significant results is small compared with a meta-analysis of outcomes [21].

Figure 1. Flow diagram of article screening and inclusion.



Data Extraction

A data extraction template (Multimedia Appendix 2) was developed for this review and piloted on 5 full-text papers. The main data elements extracted included reported factors, barriers, and facilitators to use and usage, such as retention and/or completion rate of the research study. The data were used to address the objective of the review to identify barriers and facilitators that influence user engagement.

Other extracted data were intended to document study and intervention characteristics, such as the type of technology and whether the intervention was publicly available, the target population, and the length of time that participants were able to engage with the intervention during the study.

Quality Assessment

To account for the methodological variety of studies, the quality of reporting tool by Carroll et al [22] was used to assess quality. This tool has been used earlier in systematic reviews that include qualitative and quantitative studies [23]. Using this tool, articles were assessed on 4 criteria: (1) was the study design explained, (2) was the recruitment and selection of participants explained (eg, random sampling and convenience sampling), (3) were details of the data collection method provided (eg, topic guides for interviews, number of items in a survey, use of open or closed items), and (4) were details of the analysis method provided (ie, form of analysis rather than merely reporting data were analyzed). Following the tool's guidelines, studies were considered to be *adequately reported* if a "yes" was assigned to 2 or more of these criteria.

Analysis

An inductive thematic analysis [24] was used to identify common themes among these factors. This means that no preexisting coding scheme was used; rather, codes were created based on what emerged from the data.

We used a single coder approach, in which the first author iteratively identified codes from the data and refined themes throughout the analysis. Single coder approaches are methodologically sound when they include checks on validity and reliability [25]. For our analysis, validity and reliability were assessed by reviewing a selection of codes and their corresponding text with the second and third authors and by refining the codes. This process is common in qualitative research [26]. As we used emergent coding and there was no a priori codebook, a single coder approach also allowed for consistency of coding and interpretation of codes, and this approach has been used in systematic reviews [21].

The first author began the analysis by systematically reviewing each paper. For each paper, the following sections were analyzed: abstract, results or findings, and discussion. Individual codes were created each time a factor was described that affected engagement with DMHIs.

Factors were considered a barrier or facilitator if it was explicitly defined as a facilitator or barrier by the authors of the paper and/or the description in the paper pointed to it being a barrier or facilitator. For example, “participants reported they did not use mental health apps because they had privacy concerns on what would happen with their information.” In this instance, privacy concerns are identified as barriers.

A spreadsheet was used to keep track of the emerging codes. Each spreadsheet row corresponds to a single paper. The row contains the raw text of the paper that includes the identified factors and the initial codes. These codes were iteratively reviewed and compared with the raw text they were extracted from. Codes that referred to similar concepts, such as the ability to *personalize an intervention* and *customize an intervention*, were grouped together and given more descriptive names. As an understanding of the data was developed, earlier data were revisited to refine and combine codes, revalidating the previously coded material. Finally, the final codes were grouped into broader themes (eg, the roles of *age*, *gender*, and *employment* status were grouped into the theme demographic variables).

Results

Study Characteristics

As seen in Multimedia Appendix 2, the 208 articles included in this review [27-237] reported on 2 types of user studies: (1) 69 studies were needs assessments that aimed to understand user needs and attitudes toward DMHIs without or before engaging with a specific intervention as part of a study and (2) 135 studies were evaluation studies that assessed users’ experience with a specific intervention over the course of the study. In total, 4 articles included both needs assessment and evaluation. Overall, 35 articles explored general user attitudes about DMHIs without focusing on a specific technology, whereas 173 studies focused on a specific technology (Table 1). Although all studies involved interventions for mental health, some studies focused on a particular area: 45 studies focused on depression, 22 studies focused on stress, 9 studies focused on anxiety, 6 focused on eating disorders such as bulimia nervosa, 4 studies focused on mood, and 2 studies focused on loneliness.

Table 1. Type of technology studied in included articles.

Type of technology	Values, n (%)[a]
Web based	80 (38.5)
Smartphone based	57 (27.4)
Computer based, but not web based	9 (4.3)
Mobile phone (but not a smartphone)	5 (2.4)
Wearable technology	2 (1.0)
Tablet based	2 (1.0)
Combination of technologies	18 (8.7)

[a]Not all studies mentioned a particular treatment; hence, the percentages do not add up to 100%.

Measures related to user engagement included time spent using an intervention, number of log-ins, usability, acceptability, and feasibility. The usability and acceptability of the technology were assessed using qualitative methods and standard measures, such as the survey based on the Unified Theory of Acceptance and Use of Technology [27], the Mobile Application Rating Scale [28], and the System Usability and After-Scenario Questionnaire [29]. Feasibility was defined in these studies as either completion of a program offered through the intervention or retention rate, which is the number of people who completed the research study as a proportion of the people who started the study. In total, 42 studies employed qualitative interviews to understand people’s user engagement.

Factors that influenced user engagement were assessed through surveys (72/208, 34.6%), interviews (42/208, 20.2%), focus groups (34/208, 16.3%), randomized controlled trials (23/208, 11.1%), field studies (8/208, 3.8%), workshops (3/208, 1.4%), analysis of app usage data (7/208, 3.4%), and analysis of user reviews (2/208, 1.0%), using both qualitative and quantitative methods. For example, qualitative methods gathered subjective user perceptions of what formed barriers and facilitators for

them to engage with interventions. A quantitative approach explored associations between variables, such as sociodemographic factors and intervention usage data, user satisfaction, and/or interest in using DMHIs.

The number of participants involved in these studies ranged from 6 to more than 2 million. In total, 6 studies conducted a secondary analysis of the usage data of an existing intervention or health database. For these 6 studies, the sample size was relatively large, ranging from 3158 to 2,171,325 users. Among the remaining 202 studies, the sample size ranged between 6 and 1558 users. For instance, 25% (52/208) of the studies had <18 participants, 49.5% (103/208) had <40 participants, and 75% (156/208) had <177 participants. The extent to which participants were exposed to an intervention ranged from a short demonstration before a focus group or survey to up to 1 year of usage.

Quality Assessment

All studies were assessed as *adequately reported* (Multimedia Appendix 3 [143,172]). Each study reported on the research question, study design, and method of data collection. Overall, 11 studies did not report the recruitment and/or selection process of study participants [30-40]. In addition, 11 studies did not specify the analysis method used to analyze the data [33,41-50]. One study reported on the analysis method for the quantitative data that were collected but not qualitative data [51].

Intervention Characteristics

Table 2 shows the types of technologies studied in the articles, and Table 3 shows the types of treatments and/or resources offered by the technology. Web- and smartphone-based interventions were the most common, reported in 38.5% (80/208) and 27.4% (57/208) of the papers, respectively. The most common type of treatment is internet-based CBT. Other treatments and features included acceptance and commitment therapy, psychotherapy, positive psychological interventions, meditation, peer support, resources, monitoring of symptoms, and journaling.

The target population included students, transitional age youth (aged 16-24 years), refugees, people who were homeless, veterans diagnosed with post-traumatic stress disorder, mothers with postpartum depression, patients being treated for a mental illness or another health concern, older adults, and caregivers and workers experiencing stress. Not all interventions specified the target population.

Table 2. Type of technology studied in included articles.

Type of technology	Values, n (%)[a]
Web-based	80 (38.5)
Smartphone-based	57 (27.4)
Computer-based, but not web-based	9 (4.3)
Mobile phone (but not a smartphone)	5 (2.4)
Wearable technology	2 (1.0)
Tablet-based	2 (1.0)
Combination of technologies	18 (8.7)

[a]Not all studies mentioned a particular treatment; hence, the percentages do not add up to 100%.

Table 3. Type of treatment and resources offered.

Type of treatment or resources	Values, n (%)[a]
Cognitive behavioral therapy	30 (14.4)
Informational or educational resources	23 (11.1)
Counseling	17 (8.2)
Self-tracking tools (eg, journaling, monitoring symptoms)	12 (5.8)
Mindfulness	9 (4.3)
Acceptance and commitment therapy	8 (2.9)
Peer support (eg, peer chat)	7 (3.4)
Text messaging (eg, reminders)	4 (1.9)
Positive psychology interventions	3 (1.4)
Prolonged exposure therapy	1 (0.5)
Passive data collection	1 (0.5)
Combination of treatments and/or resources	40 (19.2)

[a]Not all studies mentioned a particular treatment; hence, the percentages do not add up to 100%.

Constructs Associated With User Engagement

Textbox 4 shows the high-level constructs derived from the thematic analysis influencing user engagement with DMHIs, where the numbers in parentheses show the number of articles in which the constructs were identified. We caution that the most frequently occurring constructs are not necessarily the most *important* but rather indicate that more studies have reported on this topic. Table 4 summarizes the main findings associated with each construct. After several iterations of grouping and coding, 16 larger groups remained: demographic variables, personal traits, mental health status, beliefs, mental health and technology experience and skills, integration into life, type of content, perceived fit, perceived usefulness, level of guidance, social connectedness, impact of intervention, technology factors, privacy and confidentiality, social influence, and implementation. These themes fit into 3 categories: user-related factors, program-related factors, and factors related to the technology and implementation environment. The next section provides more detailed explanations. The full list of factors belonging to each construct is included in Multimedia Appendix 4.

Textbox 4. The constructs influencing user engagement, grouped as constructs related to the user, the program offered by the intervention, and the technology and (implementation) environment. The numbers in parentheses indicate the number of articles in which the constructs occurred.

User

- Demographic variables (31)
- Personal traits (5)
- Mental health status (59)
- Beliefs (55)
- Mental Health and Technology Experience and Skills (33)
- Integration into life (42)

Program

- Type of content (54)
- Perceived fit (61)
- Perceived usefulness (35)
- Level of guidance (40)
- Social connectedness (53)
- Impact of intervention (62)

Technology and environment

- Technology factors (100)
- Privacy and confidentiality (47)
- Social influence (16)
- Implementation (39)

Table 4. Summary of findings for each construct.

Construct	Summary of main findings
User-related constructs	
Demographic variables (sociodemographic factors, such as age, gender, and education)	Overall, women were more likely to engage with DMHIs[a] than men
Personal traits (factors related to personality traits, such as neuroticism and extraversion)	The personality traits neuroticism, agreeableness, openness, and resistance to change were associated with higher engagement, whereas extraversion was associated with lower engagement
Mental health status (factors related to the current mental health status of the user, such as the type and severity of symptoms)	Severity of mental health symptoms increased the interest in DMHIs, but symptoms related to depression, mood, and fatigue were a barrier to actual engagement
Beliefs (beliefs held by the user with regard to technology, mental health, and mental health services)	People's positive beliefs about mental health help-seeking and technology-facilitated engagement
Mental health and technology experience and skills (previous experience the user has had with technology, mental health technology, and mental health services and skills related to their digital or mental health or digital health literacy)	Digital health literacy and positive experiences with mental health services and technology were facilitators to engagement
Integration into life (the extent to which the user is able to find time and space to use the intervention and make the intervention part of their routine or life)	Engagement was facilitated if people were able to integrate DMHI use into their daily lives
Program-related constructs	
Type of content (the type of content and features offered by the intervention)	Engagement was facilitated if content was credible and if activities offered by the DMHI were of an appropriate length (ie, not too short or too long)
Perceived fit (factors related to how well the intervention is appropriate to the user's culture and values and is adaptable to the user's needs rather than a one-size-fits-all solution)	Engagement was facilitated if information offered by a DMHI was customizable and relevant to the user
Perceived usefulness (factors related to expected benefits of using the digital intervention over existing resources)	Participants were more likely to engage with DMHIs if they understood the data and knew how to use it
Level of guidance (the level of guidance offered by the intervention on how [eg, when, how often] to use it, for example, through notifications or a coach)	Guided interventions, either through a human therapist or automated reminders to use a DMHI, had higher engagement than unguided interventions
Social connectedness (the extent to which the intervention connects or isolates the user with or from others)	Being able to connect with other people through a DMHI facilitated engagement
Impact of intervention (the impact that intervention usage had on the user, such as an improvement or exacerbation of mental health symptoms [as measured by a validated survey scale])	DMHI engagement was facilitated if participants experienced a positive impact as a result of using a DMHI, such as the improvement of symptoms
Technology- and environment-related constructs	
Technology-related factors (factors related to the technology through which the intervention is offered, such as the resources and costs required to use it, usability, and technical issues experienced by the user)	Technical issues were a common barrier to engagement
Privacy and confidentiality (factors related to data security, storage, confidentiality, and privacy of the digital intervention)	Engagement was facilitated if participants had a sense that the digital platform was private and anonymous, and they could safely disclose information
Social influence (factors from the users' social environment, such as perceptions held by their peers, family, and health care provider, that influence their intention to use an intervention)	Participants were more likely to use DMHIs if people close to them, such as family and friends, thought they should use DMHIs
Implementation (factors related to the implementation of the intervention that affects use, such as the availability of user training, the phase of the user's mental health care–seeking process during which the intervention is introduced or accessed and characteristics of the health care organization supporting the DMHI)	DMHI engagement was facilitated if people were trained on how to use it

[a]DMHI: digital mental health intervention.

User-Related Constructs

User-related factors refer to factors related to the user, such as personal beliefs, skills, and experiences.

Demographic Variables

Some demographic variables were found to be associated with DMHI engagement. Studies that found an effect of gender showed that women were more likely to adopt and engage with interventions [44,52-68]. Overall, 8 studies saw an effect of age: 2 studies found that people aged ≤50 years engaged more with interventions than older adults [66,67]. These 2 studies used a relatively large sample size (1,139 and 2,171,325), and participants were exposed to the intervention for up to 1 year. A total of 6 studies found that higher engagement was seen with adults aged ≥30 years [54,57,64,65,69-71]. These studies had a smaller sample size (samples ranged from 74 to 577 people), and participants engaged with the intervention for shorter periods (up to 12 weeks). Age was also found to influence interest and expectations: users' interest in using digital therapy interventions increased with age [72], and Krause et al [58] found that older people have higher expectations of interventions.

Chudy-Onwugaje et al [73] found that age has an interacting effect with people's depression symptoms. For people aged ≤40 years, adherence increased with depressive symptoms, but there was no association between depressive symptoms and adherence in people aged >40 years. Although the reasons for this interaction were unclear from the study, the authors of the article theorize that the effect of symptoms may interact with familiarity with technology, with younger people being more comfortable using technology.

Other demographic variables associated with user engagement were as follows: (1) employment status, with people who worked full time more likely to use the intervention than people who were retired [66] or unemployed [54,68]; (2) education, with participants with higher education reporting more acceptance of interventions than people with lower education (a high school diploma or lower) [74-76]; and (3) housing situation, with people who were experiencing homelessness responding less to messages sent by a phone intervention compared with individuals with stable housing [55].

Personal Traits

Certain personality traits were associated with willingness and interest in using DMHIs. People who scored high on neuroticism and agreeableness of the Big 5 personality traits were more interested in using smartphone apps to reduce stress [77]. In a different survey reported in the same article, neuroticism was strongly linked to self-reported stress. The cooperative nature of agreeable people made it easier to accept new technology.

In addition, extraversion was a predictor of lower likelihood to prefer web-based mental health services over in-person services [72]. People who scored high on extraversion preferred to meet and connect with a doctor in person. Other personality characteristics associated with user engagement were resistance to change and openness to experience [56]. Higher openness predicted higher engagement with mindfulness and relaxation interventions. Contrary to the hypothesis made by the authors of the article that higher resistance to change would lead to resistance to adopting a new health behavior, higher resistance instead predicted higher adherence. Once people started using the intervention, a higher resistance to change facilitated commitment to continue using the intervention.

Mental Health Status

A total of 59 studies reported that people's mental health status plays a role in participants' interest in and use of a digital intervention. First, certain mental health symptoms appeared to inhibit people's motivation and/or ability to interact with an intervention. Depressive symptoms [78] and low mood [79], as measured by validated scales, have been reported as barriers for people to access and use web-based resources. Study participants reported that feeling tired also negatively affected their motivation and ability to use an intervention [44,80]. Second, the severity of these symptoms was related to engagement with digital interventions. In needs assessment studies, participants were more willing to use DMHIs if their symptoms were more severe [38,53,62,71,81,82]. However, evaluation studies have shown that more severe symptoms hamper actual engagement with digital interventions [51,56,83-101]. Depending on the type and severity of a person's mental health symptoms, studies that involved health care providers supporting digital intervention use reported that there was sometimes a need for face-to-face contact, as issues could be difficult to address remotely via a digital platform [102-104].

Beliefs

Beliefs refer to preexisting beliefs the user has about mental health help-seeking [88], their need for help [51,105-107], the acknowledgment of having mental health needs [88], and using technology for mental health treatment [38,93,108,109]. For example, preexisting beliefs of needing help for mental health needs and having a positive perception about mental health help-seeking facilitated participants' engagement with an intervention. However, even if people acknowledged a perceived need for help and were willing to seek help, engagement with a particular intervention was then affected by a person's preconceived belief about whether a digital intervention would be effective [79,104,110-112]. In 2 studies, participants did not want to use a digital intervention because technology was seen as a stimulant and distracting [113,114].

Mental Health and Technology Experience and Skills

A positive prior experience with technology [115,116], mental health services [54,74,104,117-119], and mental health technology [72,120,121] facilitated people's intention to use interventions, as well as actual engagement. A negative prior experience formed a barrier to engaging with a DMHI [74,117,122], whereas a positive experience increased participants' engagement [54,72,104,115,116,118-121].

Mental health literacy refers to knowledge about mental health symptoms and appropriate treatment options [238]. Digital literacy refers to the skills required to use technology [239]. Digital health literacy refers to the ability to use technology to find and use health resources [240]. Participants' mental health literacy [123], digital literacy [88,103,104,122,124-127], and

digital health literacy [103,127] influenced the extent to which they were able to adapt and engage with DMHIs: for each type of literacy; higher literacy was associated with higher engagement.

Integration Into Life

Users reported that their engagement was affected by the extent to which they were able to integrate an intervention into their daily lives. Barriers that limited use included that participants felt they lacked time [44,128-131] or constantly forgot to use an intervention [93,95,129], participants felt the intervention took too much time to use [132-134], and participants experienced difficulties establishing a routine of use that worked for them [130,135].

Access to a private space to access mental health resources also affected the extent to which participants could integrate an intervention into their lives. In 3 studies, participants mentioned that as opposed to going to a health care provider office, it was challenging to find a private space at home or work to use an intervention, which formed a barrier to engaging with it [136-138].

Studies have also found difficulties among users in integrating the information and tips offered by the intervention into their lives. For example, Jonathan et al [139] evaluated a smartphone app for people with serious mental illness. Participants who spent most of their day indoors without leaving their house had a hard time trying to use the tips in actual real-life scenarios.

Summary of User-Related Constructs

In summary, user engagement with DMHIs is partly influenced by factors related to the users themselves. Demographic variables such as age, gender, employment, education, and housing situation can affect user engagement. The personality traits neuroticism, agreeableness, openness, and resistance to change facilitated engagement, whereas extraversion was a barrier.

If mental health symptoms were more severe, participants were more interested in using DMHIs, but symptoms related to depression, low mood, and tiredness prevented engagement. People's beliefs about and past experiences with mental health services and technology were facilitators if these beliefs and experiences were positive, and they formed a barrier if these beliefs and experiences were negative. Participants' literacy in understanding mental health and using technology facilitated their ability to use DMHIs, and any further engagement depended on the extent to which people were able to integrate it into their daily lives.

Program-Related Constructs

The second group of constructs is related to the type of therapy or content offered through the DMHI.

Type of Content

Higher satisfaction with the type of content and features offered increased user engagement. Uncertainty about the credibility of the information, which related to the evidence base of the intervention and the source of information, was a barrier [74,88,127,140-143]. Other factors related to the modality through which content was delivered, with some participants preferring to have audio or video options in addition to text-only information [144] and whether the content was considered by users to have a supportive, nonjudgmental tone [51,145,146].

Some interventions offered programs of a fixed length or time commitment, such as a CBT program consisting of 8 weekly sessions. The length of the program as well as the length of individual sessions played a role in participants' satisfaction and their motivation to continue with the program [88,138,147-151]. In 2 studies evaluating a self-guided CBT program for 8 weeks, the length and pace of modules negatively impacted user motivation [88], and participants reported preference for more concise modules [148], although the articles did not attempt to identify an ideal module length. In other studies evaluating a CBT program that included in-person sessions with a therapist, some participants reported preference for both longer individual therapy sessions (greater than the standard 50-70 minutes) [151] and duration of treatment to maximize benefit [150].

Perceived Fit

Perceived fit refers to the extent to which users felt the intervention was appropriate and relevant to their culture and values and/or targeted to people similar to them, rather than a one-size-fits-all solution. This fit was, for example, facilitated by relevance of information to their current situation [14,46,150,152-157] and the ability to customize or personalize the intervention [30,46,83,84,122,134,135,138,158-165]. A facilitating factor was whether users were able to identify with the people presented in the intervention [166], which could be coaches or instructors, or examples of people with similar experiences. Factors that make the information relevant and in a language suitable to the user included culturally appropriate content [133,167,168], reading level suitable to the user [168], and content presented with limited jargon or technical language [169].

Perceived Usefulness

Perceived usefulness refers to the user's experience with an intervention and their perceptions of whether the intervention would be useful to them. This perception was facilitated by whether users were able to understand the data presented to them [104,117,170], whether it was clear what action they should take [129,133,154,155,166], and whether the intervention provided a clear advantage over past or current care received [103,117,121,155,171]. Identified facilitators were easier access to services that users would otherwise not have access to [103,173] and the eliminated need to travel a long distance to a health center [121].

Level of Guidance

The level of guidance refers to the extent to which users were guided to use an intervention, for example, through reminders or a web-based supporter, holding them accountable to regularly engage with the content. A facilitating factor in using DMHIs was whether the use of the intervention increased locus of control, meaning that users felt more ownership over their own health [14,84,95,124,174,175]. However, for interventions that were completely self-guided, participants experienced difficulty

engaging with them and at times neglected to use the intervention [44,95]. Participants expressed a need to receive more structured use, for example, through app reminders or a human coach checking on them on a regular basis [49,50,113,122,133,137,139,148,150,163,175-182]. In 6 studies, users stated that they would prefer if an intervention served as a complement to existing, in-person therapy rather than a replacement for in-person therapy [30,122,134,139,166,183].

Social Connectedness

The effect that an intervention had on participants' sense of social connectedness was found to facilitate user engagement. For example, being able to connect to peers or have regular contact with a personal therapist through DMHIs facilitated engagement in 18 studies [14,32,33,104,114,122,125,133,156,184-194]. In 6 studies, a noted barrier among both users and service providers was a concern about social avoidance, that is, a concern that people might use self-guided interventions in lieu of coming into a clinic in person and engaging in therapy or group sessions [78,104,122,133,138,195]. For therapy interventions, where study participants were introduced to therapists who they did not know before using the intervention, the extent to which participants could connect emotionally (therapeutic alliance) with the therapist-influenced engagement. Participants' ratings of the quality of the emotional connection were positively related to the number of log-ins, frequency of self-monitoring mood, and completion of therapy [196].

Impact of Intervention

Participants reported that the perceived changes they experienced in their mental health as a result of using an intervention affected their further engagement. Perceived symptom improvement facilitated further engagement [89,95,103,145,146,189,197-199], whereas exacerbation of symptoms negatively impacted engagement [49,93,104,200]. Other negative impacts of the intervention were also observed as barriers to ongoing user engagement. For example, in 2 studies, some information that was shared within the digital intervention was found to trigger difficult memories or emotions [150], participants were uncomfortable with exercises or information [80], or participants were exposed to negative comments by other users [150].

Another facilitating factor was whether the intervention normalized people's experiences [79,139,157,190,201], for instance, by providing examples of other people with similar experiences. Additional positive effects that could facilitate engagement were an increased insight into users' health [79,84,114,124,134,138-140,158,176,202-205], a feeling of empowerment over being in control of their health [14,32,84,93,95,104,117,122,124,170,174,175,196,206,207], improved skills [89,189,198,208-211], such as managing negative emotions, and an improvement of participants' existing relationships with others [182,205].

Summary of Program-Related Constructs

In summary, the content offered by a DMHI had to be credible and ideally offered in more than one modality. Participants engaged with DMHIs if they felt the intervention was a good fit, which could be facilitated if content was relevant, and the DMHI was customizable, culturally appropriate, and used a language that was understandable to the participant. Engagement was facilitated by participants' perception of whether a DMHI was useful, which included whether they were able to understand the data and how to use it, and whether a DMHI provided a clear advantage over resources they already had access to.

Guided DMHIs had higher engagement than unguided interventions, and participants liked being able to connect with other people, although some studies identified concerns that DMHIs could be used to avoid in-person contact. The negative and positive impacts of DMHI use could form barriers and facilitators, respectively, to further engagement.

Technology- and Environment-Related Constructs

The third group of constructs refers to factors related to the technology itself or the implementation of the technology.

Technology-Related Factors

Technology-related factors refer to factors related to the technology through which the intervention was offered. The primary barrier to engagement noted in 25 studies was users' experience of technical issues [44,50,80,92,100,103,118,129,138,155,172,179,185,195,205,208,212-220], such as a mobile app crashing and shutting down unexpectedly; in 3 studies, participants did not have the resources required to use an intervention [171,221,222]. In 7 studies, participants expressed concerns over the eventual costs associated with using an intervention [85,93,104,123,165,223,224]. Costs could be related to the need for a smartphone, having internet access, or making purchases through the app. Usability issues formed a barrier to engaging with an intervention [46,50,78,84,148-150,157,159,170,224-228]. Examples of usability issues were difficulty finding information in an intervention [78], a time-consuming process to log in to an intervention [159], and difficulty navigating within an intervention [150,157].

In addition to technical issues that formed barriers to engagement, there were also factors related to technology that facilitated the use of mental health resources and support. Facilitating factors made possible by the technology used were the flexibility of being able to access resources at any location [47,127] at any time [41,93,97,124,129,134,167,229-233] and having a temporal record of health data, such as symptoms, that users were able to track and access over time [176,180].

Privacy and Confidentiality

Privacy and confidentiality relate to how data were stored and shared and whether users felt safe and comfortable to disclose confidential information through an intervention. In 2 studies, participants were uncomfortable about their physical location being recorded [180,234], and in a study by Nicholas et al [234], participants were more comfortable with health information being recorded such as sleep and mood than personal data being recorded such as social activity and communication logs.

Accessing mental health resources via a digital platform raised concerns regarding privacy. Facilitators of user engagement and feeling safe to disclose information included assurance that the digital platform was private and participants' information could not be easily accessed by third parties [129,158,205,235].

Participants in 5 studies expressed that concerns about confidentiality formed a barrier to engagement [51,104,127,236,237]. A facilitator to create a safe environment was moderation of the intervention [140], which means that a person was monitoring and moderating the content shared by and between users within an intervention.

Anonymity was found to be both a facilitator and a barrier to engagement. Overall, 7 studies listed anonymity, meaning that users could share and receive information anonymously, as a facilitating factor to engage and encourage disclosure of information [41,88,129,137,141,148,232]. However, anonymity could also make it more difficult for participants to trust a coach who they did not know [85,127]. In these studies, participants interacted with the coach through text, and there was the option to disclose names, but neither side could see each other. Other study participants were concerned about whether an intervention was truly anonymous if it was used in a small setting, with a limited number of known users [137]. Anonymity was also more important for people who were older and who had previous experience with medical treatment [127].

Social Influence

Users' engagement was facilitated by whether the intervention was endorsed by other users [43] or peers [211,241], their friends and family [97], or their current health care provider [152,161,242]. However, if participants felt forced by others to use an intervention, it deterred them from using it [103]. If an intervention was used as part of ongoing in-person therapy, the way that therapists used or were willing to use an intervention influenced participants' engagement with an intervention [84,103,127,150,169,212,243]. The adoption of an intervention as part of therapy depended on the therapist's digital literacy skills [244,245], their past experience with mental health technology [120], and the ability to easily integrate its use into their practice as a provider [132,195,205,211,224,226,246].

Implementation

Although most studies in this review (93%, 194/208) primarily focused on factors related to the user and the intervention itself, 14 studies also described factors related to the implementation of the intervention. Examples included whether users received training on how to use the intervention [115,247] and if it was introduced early on or at a later stage in ongoing therapy. Participants in a study by Graham et al [68] used an intervention to support their mental health while in treatment for substance use. These participants found the intervention more useful at a later stage, as they felt the user was likely more familiar with their health and better able to make sense of the information provided by the intervention. Two other studies found that participants engaged more with an intervention if they were just starting treatment [104,178]. Two studies found that the way in which the intervention was labeled and introduced to users also mattered. For example, the term *mental health* was disliked by participants [248], and participants reported that they would be more likely to use an app if it was meant for *well-being* and *mental fitness* rather than mental health [37]. Other implementation factors were administrative barriers [42,118,129,211] and barriers related to the organization in which the intervention was or would be implemented [118,122,135]. Examples of administrative barriers were inadequate staffing and poor communication among staff members. An example of organizational barriers was a lack of support for DMHIs among managers.

Summary of Technology- and Environment-Related Constructs

In summary, although DMHIs introduced technical and usability issues that could form a barrier for participants to engage, the digital format also provided flexibility to access resources anywhere at any time and to have a record of health data. It was important that information was private and that participants could safely disclose information anonymously, although complete anonymity also made it more difficult to trust other people on the platform. Negative and positive opinions held by other people about DMHIs could form a barrier and facilitator, respectively, to engagement, and if DMHIs were to be used as part of ongoing therapy, the therapists' past experience with DMHIs and the ability to integrate it into their practice played a role in user engagement. Finally, successful implementation facilitated user engagement. Providing training on how to use DMHIs and labeling an intervention for well-being or mental fitness (as opposed to mental health) can help users engage with DMHIs more. Participants may be more engaged with DMHIs if they are just starting treatment, but the identified benefit of introducing DMHIs at a later stage is that users may be more knowledgeable about their health and better able to make sense of their health information.

Discussion

Principal Findings

This study aims to synthesize the literature on DMHIs and summarize the identified factors affecting user engagement with DMHIs. This review identifies 3 key areas that all contribute to DMHI engagement: (1) user characteristics, such as severe mental health symptoms, can form a barrier to engagement; (2) users' experience of the program or content, with participants more likely to engage if they perceive the program to be useful and a good fit to them; and (3) the technology and implementation environment, such as technical issues being a common barrier to engaging with DMHIs. Providing content that is relevant and customizable according to personal preferences and offering technical assistance and/or training are important to achieve engagement. However, although these considerations may increase interest and uptake of DMHIs, it is important to understand whether characteristics specific to the user, such as their symptoms, will affect motivation to engage with these interventions. We first discuss the 3 key areas in more detail in the following three subsections; compare our constructs with other models on user engagement; and then discuss implications for researchers, developers, and health service providers.

User Constructs

Individual differences among users can affect engagement, including demographic variables such as age and gender, personality traits, mental health status, beliefs about mental health and DMHIs, experience with technology and mental

health, and people's ability to integrate DMHI use into their lives. Although the severity of symptoms may increase interest in engaging with health interventions [249], symptoms related to depression, mood, and tiredness were found to hamper actual engagement. This contrast may point to the unique implications that mental health symptoms can have on engagement with DMHIs.

The contrasting role of symptom severity between studies highlights the importance of understanding how people who would be more interested in DMHIs and may benefit more from its use are not limited by their symptoms to actually engage with these interventions. The contrast also illustrates the importance of including users at various stages of the design process, as people may be interested in the concept of a DMHI but may not be able to actually engage with it because of the nature of their symptoms.

Although studies looking at DMHI usage over 1 year found that younger people were more engaged with DMHIs, shorter research studies (ie, up to 12 weeks) found that older people were more engaged. Potentially, older adults perform better on study adherence, and younger people continue to engage more with an intervention long term, although the different interventions and settings make it difficult to make a direct comparison between these studies.

Program Constructs

Engagement with DMHIs was facilitated if participants liked the type of content; they perceived a DMHI to be a good fit for them and perceived it to be useful; there was a level of guidance on how to use it, it facilitated social connectedness, and it had a positive impact, such as improvement of symptoms.

Level of Guidance

Guided interventions typically have higher engagement than unguided interventions. However, human guidance can be resource intensive, and it may not always be possible or feasible to provide the desired level of guidance. Although human support enhances engagement more than automated means such as email reminders [250], several studies included in our review found that such automated reminders not only facilitated engagement but were also experienced positively by users. Automated reminders to use an intervention may therefore be a low-cost alternative to human support. The benefits of automated reminders may depend on the type of support and the type of barriers they are designed to address. Short text-based reminders may be suitable for in-the-moment interventions [251] and may be useful to address barriers to forgetting to use an intervention. On the other hand, human support may be more suitable for addressing the lack of motivation and facilitating social connectedness.

Furthermore, appropriate time commitments differ for self-guided exercises versus guided sessions. Participants across studies preferred shorter self-guided modules but longer guided therapy sessions. Finally, personalization may also meet different preferences. People who find videos or text-based material time consuming may be more engaged with shorter actionable exercises, whereas people with a preference for synchronous communication may engage more when they get dedicated time on one-on-one sessions. It would be worthwhile to further explore how engagement can be encouraged in self-guided interventions.

Social Connectedness

An important facilitator was whether a DMHI facilitated social connectedness and enabled the user to interact with other people. Previous work has shown that social support through social networks not only increases engagement but may also have a positive effect on depression symptoms [221,222]. However, in some studies in this review, mental health service users and providers were concerned that technology would facilitate social avoidance if people were to use a digital intervention in lieu of engaging in face-to-face individual or group therapy. It appears that it is important that an intervention allows users to connect with other people with whom they may have otherwise not connected, rather than replacing any existing face-to-face contact. For example, people can access a mental health app if they are not able to speak to someone in person about their concerns [175].

Technology and Environment Constructs

Offering mental health resources through technology offers both barriers and facilitators. Technical issues and concerns about privacy were common barriers, but technology also offered flexibility and could facilitate anonymity. Furthermore, the environmental context in which DMHIs are to be used are important to consider. Participants were more likely to use DMHIs if people close to them thought they should use it and if they received training on how to use it.

Anonymity

Anonymity was a prominent topic among studies but engaging with an intervention anonymously was seen as both a barrier to and facilitator of engagement, sometimes within the same study. This difference can be explained by factors related to the user, the type of implementation setting, and the type of intervention features that were anonymous, as outlined in the following paragraph.

First, a facilitating aspect of an anonymous intervention was that study participants found it less stigmatizing than seeing a live or in-person therapist. Anonymity may be an important facilitator for people who have experienced stigma and embarrassment, which is known to be a barrier to help-seeking for mental health concerns [2,3]. Similarly, prior work on mental health discourse on the web found that anonymity does not hinder the social support that people receive on their posts, which can facilitate open conversations, and that social media may be particularly useful for stigmatic illnesses such as mental health [252]. Second, the study setting matters. Interventions that are used in a relatively small setting may give a false sense of anonymity if it is possible for users to find out who else is using the intervention, for example, through content shared within the intervention or by seeing someone use it [108], which is important to consider for intimate settings, such as schools, workplaces, or small communities. Third, on community forums, where users could share their experiences and comment on other users' posts, overall anonymity was seen as a facilitator to safely disclose information. In one-on-one sessions, however, where

the user interacted with a coach or therapist but neither side could see each other, anonymity made it more difficult to establish a relationship and trust compared with a face-to-face session.

These differing perceptions shed light on an important trade-off. Should an intervention strive to be anonymous to address stigma and potential embarrassment or focus on allowing people to establish a trusted relationship with someone? This decision may depend on the objective of the intervention and whether anonymity is possible in the context in which it is to be used. Alternatively, a hybrid form or multiple options can be considered and offered. For example, forums with a larger number of users can be anonymous, whereas a private one-on-one session with a therapist can include telehealth options to allow for therapeutic alliance building between the user and therapist. The Supportive Accountability Model [250] also proposes that engagement is enhanced if human coaches are seen as trustworthy and that users may disclose more in computer-mediated than face-to-face communication. Although Mohr et al [250] argue that providing additional information about individuals, such as photographs, may reduce these types of benefits of this type of mediated communication, it may be important to establish initial trust with a coach or therapist. Additional research is needed to understand how best to support trust in DMHIs.

Privacy

A previous review of user engagement with mental health apps theorized that one reason for low engagement is that these apps do not consider user privacy [15]. In our review, privacy was discussed in terms of data storage and sharing but also with respect to the physical environment in which these interventions were to be used. Delivering mental health support through a digital platform was found to increase a sense of privacy in some studies, but in other studies, it was associated with a decreased sense of privacy. The study participants stated that they could access care more privately, without anyone knowing about it. In line with previous work [2], this again indicates that privacy can be important for people who experience stigma for or reluctance to help-seeking. Although participants' living situation was not explicitly discussed in these studies, when compared with other studies, it is likely that participants were able to engage with these interventions in physically private settings. In other studies where people did not have access to a private space, a lack of privacy was a barrier to engagement. For example, study participants evaluating an app that delivered remote web-based therapy felt that they could disclose more in a closed therapist office than through a web-based intervention at home where other people in their household could see or disrupt them [136]. Study participants who used a mental health intervention in the workplace [137] said privacy was not possible, as colleagues could see what someone was doing at their desk and when they were interacting with the intervention.

These differing experiences highlight that technology can overcome existing privacy barriers of seeking mental health care but can also introduce other privacy issues, and users' situational context (ie, where they are physically accessing the digital intervention) should be taken into account.

Comparison With Other Models of Technology and Digital Health Intervention Engagement

Some of the themes identified in this review overlap with previous models conceptualizing engagement with digital health interventions, as well as general technology acceptance and health behavior, such as the Efficiency Model of Support [251], Technology Acceptance Model [253], and Health Belief Model [249].

For instance, the Efficiency Model of Support [251] states that human support increases engagement in the context of the use of digital health interventions when it addresses 1 of 5 failure points: *usability*, *implementation*, *fit*, *engagement*, and *knowledge*. These broadly map to our constructs of technology factors, integration into life, perceived fit, beliefs, and experience, and skills. The *implementation* failure point in the efficiency model pertained to whether the user can apply knowledge gained from an intervention in their lives. Our review extends this concept, in that we found that an important issue is whether users can integrate the actual use of the intervention into their everyday routines.

Our findings are in line with the Technology Acceptance Model, which explains that users' decisions to accept and use a technology are influenced by perceived usefulness, ease of use, and social influence of others. The Health Belief Model explains that adoption of health interventions is, among things, influenced by a person's belief in the severity of their illness or health symptoms and the perceived benefits of seeking treatment for these symptoms, which map onto our constructs of beliefs and impact of the intervention. Themes revealed in this review, which have not been highlighted in these previous models, are the level of guidance, integration into life, and social connectedness. This gap may be explained by the way in which mental health interventions were intended to be used. To be effective, most DMHIs were intended to be used regularly by users on their own. This characteristic introduces the challenge for people to integrate it into their routine and have the discipline to use it regularly; therefore, the level of guidance provided within the intervention may have a particularly salient effect on engagement. Social connectedness may be especially important for mental health interventions, as it can improve mood [254] and help combat depression [255].

Implications

In this review, we have synthesized the literature on DMHIs to identify common factors influencing user engagement. This synthesis can be described as follows.

- Researchers can use these factors to develop constructs that are important to measure when evaluating DMHIs. More concretely, it is important to capture user characteristics, users' experience of the program and content, and details regarding the implementation setting. These constructs may help explain why someone would use one DMHI over another and may help evaluate how engaging a DMHI will be.
- Developers can use these factors to facilitate engagement with DMHIs. Specifically, when developing a DMHI, it is important to understand the specific characteristics of the

target audience, for example, if the severity of the audience's symptoms can form a barrier to engagement; to tailor the program to the audience, such as offering the option to customize content; and to address issues related to the technology and environment, for example, by mitigating technical issues and providing technical assistance.

- Mental health service providers, such as clinicians, can use this overview as guidance to select interventions that are appropriate for their clients or help guide their clients in selecting suitable interventions. For example, it is important to consider whether an intervention can be easily integrated into clients' lives and routines. In addition,Multimedia Appendix 2, which shows the full data set, can be used to filter the study setting, target population, and symptoms to see which barriers and facilitators have been observed for similar settings and populations.

The themes highlighted in this review identify factors that can facilitate engagement and barriers that should be considered to facilitate the successful implementation of a digitally mediated mental health intervention.

Limitations

We did not limit this review to particular study designs. As such, this review takes a much broader look at what factors influence engagement with digital mental health technologies rather than focusing on a single research method or technology. However, because of the heterogeneity of the included studies, we were unable to conduct a meta-analysis. In addition, there was inconsistency across studies in measures used to assess user engagement, such as the number of log-ins to an intervention, the length of continuing to engage with it, the total time spent using an intervention, or a self-reported measure of engagement by participants. This inconsistency has been found to be an issue in previous reviews on the user engagement of DMHIs [11,256]. This review was limited to peer-reviewed empirical articles. Although the review included articles that evaluated people's experience with both research interventions and commercially available DMHIs, it is possible that some interventions may have been missed.

Finally, this review was conducted before the global COVID-19 pandemic. There may be unique factors that are pandemic related that make DMHI engagement more or less likely. For example, stay-at-home orders may exacerbate feelings of social isolation and make people more likely to engage with apps that increase social connectedness. On the other hand, it may also introduce additional barriers to finding a private space to use DMHIs if sheltering in place with others. The results presented in this review should be interpreted and used to understand DMHI engagement before and after the pandemic. A future review could be conducted solely during the pandemic period, and it could be compared with this review to understand DMHI use outside versus during a pandemic.

Conclusions

Previous studies have shown the potential of DMHIs to improve mental health. However, for these interventions to be clinically effective, they require engagement by users in real-world settings. Across the studies reviewed, we identified 16 common factors that affect user engagement. Further research on DMHIs can use these factors as guidelines when evaluating interventions with users, and future interventions can be developed with these factors in mind. By understanding the factors that affect engagement, targeted strategies can be developed to overcome addressable barriers and work toward the successful implementation of these interventions.

Acknowledgments

The authors thank Vicky Yu who helped with data entry for this manuscript.

Conflicts of Interest

SMS has received consulting payments from Otsuka Pharmaceuticals. All other authors declared no conflicts of interest.

Multimedia Appendix 1
Overview of search queries or terms used for each database.
[DOC File , 17 KB - jmir_v23i3e24387_app1.doc]

Multimedia Appendix 2
Data extraction template with metadata of the reviewed articles.
[XLS File (Microsoft Excel File), 84 KB - jmir_v23i3e24387_app2.xls]

Multimedia Appendix 3
Quality assessment of included studies.
[DOCX File , 41 KB - jmir_v23i3e24387_app3.docx]

Multimedia Appendix 4
Overview of barriers and facilitators for each theme.
[XLS File (Microsoft Excel File), 12 KB - jmir_v23i3e24387_app4.xls]

References

1. Lipari, RN, Park-Lee E. Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health Internet. Substance Abuse and Mental Health Services Administration. 2019. URL: https://www.samhsa.gov/data/ [accessed 2020-07-01]
2. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10(113):1-9. [doi: 10.1186/1471-244x-10-113]
3. America's Mental Health 2018. Cohen Veterans Network. 2018. URL: https://www.cohenveteransnetwork.org/wp-content/uploads/2018/10/Research-Summary-10-10-2018.pdf [accessed 2020-07-01]
4. Anderson, NB, Belar CD, Breckler SJ, Nordal KC, Ballard DW, Bufka LF, et al. Stress in America: Paying With Our Health. American Psychological Association. 2014. URL: https://www.apa.org/news/press/releases/stress/2014/stress-report.pdf [accessed 2020-07-01]
5. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Aug;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
6. Muñoz RF. The Efficiency Model of Support and the Creation of Digital Apothecaries. Clin Psychol Sci Pract 2016 Nov 17;24(1):46-49. [doi: 10.1111/cpsp.12174]
7. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of Self-guided Internet-Based Cognitive Behavioral Therapy in the Treatment of Depressive Symptoms: A Meta-analysis of Individual Participant Data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
8. Graham AK, Greene CJ, Kwasny MJ, Kaiser SM, Lieponis P, Powell T, et al. Coached Mobile App Platform for the Treatment of Depression and Anxiety Among Primary Care Patients: A Randomized Clinical Trial. JAMA Psychiatry 2020 Sep 01;77(9):906-914. [doi: 10.1001/jamapsychiatry.2020.1011] [Medline: 32432695]
9. Baumel A, Edan S, Kane J. Is there a trial bias impacting user engagement with unguided e-mental health interventions? A systematic comparison of published reports and real-world usage of the same programs. Transl Behav Med 2019 Nov 25;9(6):1020-1033 [FREE Full text] [doi: 10.1093/tbm/ibz147] [Medline: 31689344]
10. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
11. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the Trial: Systematic Review of Real-World Uptake and Engagement With Digital Self-Help Interventions for Depression, Low Mood, or Anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
12. Kaltenthaler E, Sutcliffe P, Parry G, Beverley C, Rees A, Ferriter M. The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. Psychol Med 2008 Nov;38(11):1521-1530. [doi: 10.1017/S0033291707002607] [Medline: 18205964]
13. Feng XL, Campbell A. Understanding E-Mental Health Resources: Personality, Awareness, Utilization, and Effectiveness of E-Mental Health Resources Amongst Youth. Journal of Technology in Human Services 2011 Apr;29(2):101-119. [doi: 10.1080/15228835.2011.595276]
14. Wilhelmsen M, Lillevoll K, Risør M, Høifødt R, Johansen M, Waterloo K, et al. Motivation to persist with internet-based cognitive behavioural treatment using blended care: a qualitative study. BMC Psychiatry 2013;13(1):296. [doi: 10.1186/1471-244x-13-296]
15. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Mental Health 2018 Jun 05;21(3):116-119. [doi: 10.1136/eb-2018-102891]
16. Wasil AR, Gillespie S, Shingleton R, Wilks CR, Weisz JR. Examining the Reach of Smartphone Apps for Depression and Anxiety. AJP 2020 May 01;177(5):464-465. [doi: 10.1176/appi.ajp.2019.19090905]
17. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci. Transl. Med 2015 Apr 15;7(283):283rv3-283rv3. [doi: 10.1126/scitranslmed.aaa3487]
18. Batra S, Baker R, Wang T, Forma F, DiBiasi F, Peters-Strickland T. Digital health technology for use in patients with serious mental illness: a systematic review of the literature. MDER 2017 Oct;10:237-251. [doi: 10.2147/mder.s144158]
19. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital Mental Health Interventions for Depression, Anxiety, and Enhancement of Psychological Well-Being Among College Students: Systematic Review. J Med Internet Res 2019 Jul 22;21(7):e12869. [doi: 10.2196/12869]
20. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(210):1-10. [doi: 10.1186/s13643-016-0384-4]
21. Lienemann BA, Unger JB, Cruz TB, Chu K. Methods for Coding Tobacco-Related Twitter Data: A Systematic Review. J Med Internet Res 2017 Mar 31;19(3):e91 [FREE Full text] [doi: 10.2196/jmir.7022] [Medline: 28363883]
22. Carroll C, Booth A, Lloyd-Jones M. Should We Exclude Inadequately Reported Studies From Qualitative Systematic Reviews? An Evaluation of Sensitivity Analyses in Two Case Study Reviews. Qual Health Res Internet 2012;22(10):1425-1434 [FREE Full text] [doi: 10.1177/1049732312452937]

23. Golder S, Ahmed S, Norman G, Booth A. Attitudes Toward the Ethics of Research Using Social Media: A Systematic Review. J Med Internet Res 2017 Jun 06;19(6):e195. [doi: 10.2196/jmir.7082]
24. Braun A, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Aug;3(2):77-101. [doi: 10.1002/0471142905.hg1201s50] [Medline: 18428388]
25. Ng A, Reddy M, Zalta A, Schueller SM. Veterans' perspectives on fitbit use in treatment for post-traumatic stress disorder: An interview study. J Med Internet Res 2018;20(6):e10415. [doi: 10.2196/10415]
26. Guest G, MacQueen K, Namey E. Applied Thematic Analysis. Thousand Oaks, CA: SAGE Publications; 2011.
27. Venkatesh V, Morris M, Davis G, Davis F. User Acceptance of Information Technology: Toward a Unified View. MIS Q Internet 2003;27(3):425-478 [FREE Full text] [doi: 10.2307/30036540]
28. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
29. Lewis JR. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
30. Pagliari C, Burton C, McKinstry B, Szentatotai A, David D, Serrano Blanco A, et al. Psychosocial implications of avatar use in supporting therapy for depression. Stud Health Technol Inform 2012;181:329-333. [Medline: 22954882]
31. Moessner M, Minarik C, Özer F, Bauer S. Can an internet-based program for the prevention and early intervention in eating disorders facilitate access to conventional professional healthcare? Journal of Mental Health 2016 Feb 05;25(5):441-447. [doi: 10.3109/09638237.2016.1139064]
32. Kenny R, Dooley B, Fitzgerald A. Developing mental health mobile apps: Exploring adolescents' perspectives. Health Informatics J 2014 Nov 10;22(2):265-275. [doi: 10.1177/1460458214555041] [Medline: 25385165]
33. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JEWC. Mixing Online and Face-to-Face Therapy: How to Benefit From Blended Care in Mental Health Care. JMIR Ment Health 2016;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
34. Schlosser D, Campellone T, Kim D, Truong B, Vergani S, Ward C, et al. Feasibility of PRIME: A Cognitive Neuroscience-Informed Mobile App Intervention to Enhance Motivated Behavior and Improve Quality of Life in Recent Onset Schizophrenia. JMIR Res Protoc 2016 Apr 28;5(2):e77. [doi: 10.2196/resprot.5450]
35. Puszka S, Dingwall KM, Sweet M, Nagel T. E-Mental Health Innovations for Aboriginal and Torres Strait Islander Australians: A Qualitative Study of Implementation Needs in Health Services. JMIR Ment Health 2016 Sep 19;3(3):e43 [FREE Full text] [doi: 10.2196/mental.5837] [Medline: 27644259]
36. Bennett-Levy J, Singer J, DuBois S, Hyde K. Translating E-Mental Health Into Practice: What Are the Barriers and Enablers to E-Mental Health Implementation by Aboriginal and Torres Strait Islander Health Professionals? J Med Internet Res 2017 Jan 11;19(1):e1 [FREE Full text] [doi: 10.2196/jmir.6269] [Medline: 28077347]
37. Deady M, Peters D, Lang H, Calvo R, Glozier N, Christensen H, et al. Designing smartphone mental health applications for emergency service workers. Occup Med (Lond) 2017 Aug 01;67(6):425-428. [doi: 10.1093/occmed/kqx056] [Medline: 28535246]
38. Arjadi R, Nauta MH, Bockting CLH. Acceptability of internet-based interventions for depression in Indonesia. Internet Interv 2018 Sep 01;13(2):8-15 [FREE Full text] [doi: 10.1016/j.invent.2018.04.004] [Medline: 30206513]
39. Stallman HM, Kavanagh DJ. Development of an Internet Intervention to Promote Wellbeing in College Students. Aust Psychol 2016 Aug 28;53(1):60-67. [doi: 10.1111/ap.12246]
40. Clough BA, Zarean M, Ruane I, Mateo NJ, Aliyeva TA, Casey LM. Going global: do consumer preferences, attitudes, and barriers to using e-mental health services differ across countries? Journal of Mental Health 2017 Aug 31;28(1):17-25. [doi: 10.1080/09638237.2017.1370639]
41. Gulec H, Moessner M, Mezei A, Kohls E, Túry F, Bauer S. Internet-based maintenance treatment for patients with eating disorders. Professional Psychology: Research and Practice 2011;42(6):479-486. [doi: 10.1037/a0025806]
42. Adler G, Pritchett LR, Kauth MR, Nadorff D. A pilot project to improve access to telepsychotherapy at rural clinics. Telemed J E Health 2014 Jan;20(1):83-85 [FREE Full text] [doi: 10.1089/tmj.2013.0085] [Medline: 24168721]
43. Batterham P, Calear A, O'Dea B, Larsen M, Titov N, March S, et al. Stakeholder perspectives on evidence for digital mental health interventions: Implications for accreditation systems. Digit Health 2019;5:1-7. [doi: 10.1177/2055207619878069]
44. Harjumaa M, Halttu K, Koistinen K, Oinas-Kukkonen H. User Experience of Mobile Coaching for Stress-Management to Tackle Prevalent Health Complaints. : Springer International Publishing; 2015 Presented at: Scandinavian Conference on Information Systems; August 9-12; Oulu, Finland p. 152-164. [doi: 10.1007/978-3-319-21783-3_11]
45. Rickard N, Arjmand H, Bakker D, Seabrook E. Development of a Mobile Phone App to Support Self-Monitoring of Emotional Well-Being: A Mental Health Digital Innovation. JMIR Ment Health 2016 Nov 23;3(4):e49 [FREE Full text] [doi: 10.2196/mental.6202] [Medline: 27881358]
46. Rodriguez-Paras C, Sasangohar F. Usability Assessment of a Post-Traumatic Stress Disorder (PTSD) mHealth App. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2017 Oct 20 Presented at: Proc Hum Factors Ergon Soc; 2017; Austin, Texas p. 1824-1828. [doi: 10.1177/1541931213601937]

47. Melton L, Brewer B, Kolva E, Joshi T, Bunch M. Increasing access to care for young adults with cancer: Results of a quality-improvement project using a novel telemedicine approach to supportive group psychotherapy. Palliat Support Care 2017 Apr;15(2):176-180. [doi: 10.1017/S1478951516000572] [Medline: 27457109]
48. Simões de Almeida R, Sousa T, Marques A, Queirós C. Patients' perspectives about the design of a mobile application for psychotic disorders. Psychol. Community Health 2018 Aug 03;7(1):16-28. [doi: 10.5964/pch.v7i1.192]
49. Simões de Almeida R, Sousa T, Couto A, Marques A, Queirós C, Martins C. Development of weCope, a mobile app for illness self-management in schizophrenia. Clinical Psychiatry 2019;46(1):1-4. [doi: 10.1590/0101-60830000000182]
50. Thach K. A Qualitative Analysis of User Reviews on Mental Health Apps: Who Used it? for What? and Why? A qualitative analysis of user reviews on mental health apps: Who used it? For what? And why? Tra Vinh University, School of Engineering and Technology, Tra Vinh, Viet Nam: Institute of Electrical and Electronics Engineers Inc; 2019 Presented at: 2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF); March 20-22; Danang, Vietnam p. 1-4. [doi: 10.1109/rivf.2019.8713726]
51. Doherty G, Coyle D, Sharry J. Engagement with online mental health interventions: An exploratory clinical study of a treatment for depression. 2012 Presented at: CHI'12; May 5-10; Austin, Texas p. 1421-1430. [doi: 10.1145/2207676.2208602]
52. Proudfoot J, Parker G, Manicavasagar V, Hadzi-Pavlovic D, Whitton A, Nicholas J, et al. Effects of adjunctive peer support on perceptions of illness control and understanding in an online psychoeducation program for bipolar disorder: A randomised controlled trial. Journal of Affective Disorders 2012 Dec;142(1-3):98-105. [doi: 10.1016/j.jad.2012.04.007]
53. Crisp DA, Griffiths KM. Participating in Online Mental Health Interventions: Who Is Most Likely to Sign Up and Why? Depression Research and Treatment 2014;2014:1-11. [doi: 10.1155/2014/790457]
54. Gunn J, Cameron J, Densley K, Davidson S, Fletcher S, Palmer V, et al. Uptake of mental health websites in primary care: Insights from an Australian longitudinal cohort study of depression. Patient Educ Couns 2018;101(1):105-112 [FREE Full text] [doi: 10.1016/j.pec.2017.07.008]
55. Mackesy-Amiti ME, Boodram B. Feasibility of ecological momentary assessment to study mood and risk behavior among young people who inject drugs. Drug and Alcohol Dependence 2018 Jun;187:227-235. [doi: 10.1016/j.drugalcdep.2018.03.016]
56. Mikolasek M, Witt CM, Barth J. Adherence to a Mindfulness and Relaxation Self-Care App for Cancer Patients: Mixed-Methods Feasibility Study. JMIR Mhealth Uhealth 2018 Dec 06;6(12):e11271 [FREE Full text] [doi: 10.2196/11271] [Medline: 30522990]
57. Achtyes ED, Ben-Zeev D, Luo Z, Mayle H, Burke B, Rotondi AJ, et al. Off-hours use of a smartphone intervention to extend support for individuals with schizophrenia spectrum disorders recently discharged from a psychiatric hospital. Schizophrenia Research 2019 Apr;206:200-208. [doi: 10.1016/j.schres.2018.11.026]
58. Krause K, Guertler D, Moehring A, Batra A, Eck S, Rumpf H, et al. Feasibility and Acceptability of an Intervention Providing Computer-Generated Tailored Feedback to Target Alcohol Consumption and Depressive Symptoms in Proactively Recruited Health Care Patients and Reactively Recruited Media Volunteers: Results of a Pilot Study. Eur Addict Res 2019 Dec 15;25(3):119-131. [doi: 10.1159/000499040] [Medline: 30917380]
59. La Porte L, Kim J, Adams M, Zagorsky B, Gibbons R, Silver R, et al. Feasibility of perinatal mood screeningtext messaging on patients? personal smartphones. Arch Womens Ment Health 2019;23(2):181-188. [doi: 10.1007/s00737-019-00981-5]
60. Mira A, Soler C, Alda M, Baños R, Castilla D, Castro A, et al. Exploring the Relationship Between the Acceptability of an Internet-Based Intervention for Depression in Primary Care and Clinical Outcomes: Secondary Analysis of a Randomized Controlled Trial. Front Psychiatry 2019;10:325 [FREE Full text] [doi: 10.3389/fpsyt.2019.00325] [Medline: 31133899]
61. Smail-Crevier R, Powers G, Noel C, Wang J. Health-Related Internet Usage and Design Feature Preference for E-Mental Health Programs Among Men and Women. J Med Internet Res 2019 Mar 18;21(3):e11224 [FREE Full text] [doi: 10.2196/11224] [Medline: 30882361]
62. Crosier B, Brian R, Ben-Zeev D. Using Facebook to reach people who experience auditory hallucinations. J Med Internet Res 2016;18(6):e160 [FREE Full text] [doi: 10.2196/jmir.5420]
63. Kenny R, Dooley B, Fitzgerald A. Ecological Momentary Assessment of Adolescent Problems, Coping Efficacy, and Mood States Using a Mobile Phone App: An Exploratory Study. JMIR Ment Health 2016 Nov 29;3(4):e51. [doi: 10.2196/mental.6361]
64. Mattila E, Lappalainen R, Välkkynen P, Sairanen E, Lappalainen P, Karhunen L, et al. Usage and Dose Response of a Mobile Acceptance and Commitment Therapy App: Secondary Analysis of the Intervention Arm of a Randomized Controlled Trial. JMIR Mhealth Uhealth 2016 Jul 28;4(3):e90. [doi: 10.2196/mhealth.5241]
65. Beatty L, Kemp E, Binnion C, Turner J, Milne D, Butow P, et al. Uptake and adherence to an online intervention for cancer-related distress: older age is not a barrier to adherence but may be a barrier to uptake. Support Care Cancer 2017 Jun;25(6):1905-1914. [doi: 10.1007/s00520-017-3591-1] [Medline: 28155018]
66. Kannisto KA, Korhonen J, Adams CE, Koivunen MH, Vahlberg T, Välimäki MA. Factors Associated With Dropout During Recruitment and Follow-Up Periods of a mHealth-Based Randomized Controlled Trial for Mobile.Net to Encourage Treatment Adherence for People With Serious Mental Health Problems. J Med Internet Res 2017 Feb 21;19(2):e46. [doi: 10.2196/jmir.6417] [Medline: 28223262]

67. Abel EA, Shimada SL, Wang K, Ramsey C, Skanderson M, Erdos J, et al. Dual Use of a Patient Portal and Clinical Video Telehealth by Veterans with Mental Health Diagnoses: Retrospective, Cross-Sectional Analysis. J Med Internet Res 2018 Nov 07;20(11):e11350. [doi: 10.2196/11350]
68. Graham S, Irving J, Cano I, Edwards M. Participation with online recovery specific groups - findings from the UK Life in Recovery survey 2015. Alcoholism Treatment Quarterly 2018 Jul 31;36(4):459-481. [doi: 10.1080/07347324.2018.1500873]
69. Pavliscsak H, Little JR, Poropatich RK, McVeigh FL, Tong J, Tillman JS, et al. Assessment of patient engagement with a mobile application among service members in transition. J Am Med Inform Assoc 2015 Sep 11;23(1):110-118. [doi: 10.1093/jamia/ocv121]
70. Schneider BC, Schröder J, Berger T, Hohagen F, Meyer B, Späth C, et al. Bridging the "digital divide": A comparison of use and effectiveness of an online intervention for depression between Baby Boomers and Millennials. J Affect Disord 2018 Aug 15;236:243-251. [doi: 10.1016/j.jad.2018.04.101] [Medline: 29751239]
71. Pruitt LD, Vuletic S, Smolenski DJ, Wagner A, Luxton DD, Gahm GA. Predicting post treatment client satisfaction between behavioural activation for depression delivered either in-person or via home-based telehealth. J Telemed Telecare 2019 Sep;25(8):460-467. [doi: 10.1177/1357633X18784103] [Medline: 29976097]
72. March S, Day J, Ritchie G, Rowe A, Gough J, Hall T, et al. Attitudes Toward e-Mental Health Services in a Community Sample of Adults: Online Survey. J Med Internet Res 2018 Feb 19;20(2):e59 [FREE Full text] [doi: 10.2196/jmir.9109] [Medline: 29459357]
73. Chudy-Onwugaje K, Abutaleb A, Buchwald A, Langenberg P, Regueiro M, Schwartz D, et al. Age Modifies the Association Between Depressive Symptoms and Adherence to Self-Testing With Telemedicine in Patients With Inflammatory Bowel Disease. Inflamm Bowel Dis 2018;24(12):2648-2654. [doi: 10.1093/ibd/izy194]
74. Watson H, Levine M, Zerwas S, Hamer R, Crosby R, Sprecher C, et al. Predictors of dropout in face-to-face and internet-based cognitive-behavioral therapy for bulimia nervosa in a randomized controlled trial. Int J Eat Disord 2017;50(5):569-577. [doi: 10.1002/eat.22644]
75. Stevens R, Polk K, Merrill C, Feng F, Weiss M, Brosnan E, et al. User experience design to enhance the effectiveness of mobile technologies for the treatment of mental health. 2018 Presented at: 2018 Systems and Information Engineering Design Symposium (SIEDS); April 27; Charlottesville, Virginia p. 135-140. [doi: 10.1109/sieds.2018.8374724]
76. Kemmeren LL, van Schaik A, Smit JH, Ruwaard J, Rocha A, Henriques M, et al. Unraveling the Black Box: Exploring Usage Patterns of a Blended Treatment for Depression in a Multicenter Study. JMIR Ment Health 2019 Jul 25;6(7):e12707. [doi: 10.2196/12707]
77. Ervasti M, Kallio J, Määttänen I, Mäntyjärvi J, Jokela M. Influence of Personality and Differences in Stress Processing Among Finnish Students on Interest to Use a Mobile Stress Management App: Survey Study. JMIR Ment Health 2019 May 13;6(5):e10039. [doi: 10.2196/10039]
78. Crooks RE, Bell M, Patten SB, Wiebe S, Holroyd-Leduc J, Bulloch AG, et al. Mind the gap: Exploring information gaps for the development of an online resource hub for epilepsy and depression. Epilepsy Behav 2017 May;70(Pt A):18-23. [doi: 10.1016/j.yebeh.2017.03.005] [Medline: 28407525]
79. Eisner E, Drake RJ, Berry N, Barrowclough C, Emsley R, Machin M, et al. Development and Long-Term Acceptability of ExPRESS, a Mobile Phone App to Monitor Basic Symptoms and Early Signs of Psychosis Relapse. JMIR Mhealth Uhealth 2019 Mar 29;7(3):e11568. [doi: 10.2196/11568]
80. Mitchell A, Whittingham K, Steindl S, Kirby J. Feasibility and acceptability of a brief online self-compassion intervention for mothers of infants. Arch Womens Ment Health 2018;21(5):553-561. [doi: 10.1007/s00737-018-0829-y]
81. Toscos T, Carpenter M, Drouin M, Roebuck A, Kerrigan C, Mirro M. College Students' Experiences with, and Willingness to Use, Different Types of Telemental Health Resources: Do Gender, Depression/Anxiety, or Stress Levels Matter? Telemedicine and e-Health 2018 Dec;24(12):998-1005. [doi: 10.1089/tmj.2017.0243]
82. Huberty J, Vranceanu A, Carney C, Breus M, Gordon M, Puzia ME. Characteristics and Usage Patterns Among 12,151 Paid Subscribers of the Calm Meditation App: Cross-Sectional Survey. JMIR Mhealth Uhealth 2019 Nov 3;7(11):e15648. [doi: 10.2196/15648]
83. Nicholas J, Proudfoot J, Parker G, Gillis I, Burckhardt R, Manicavasagar V, et al. The Ins and Outs of an Online Bipolar Education Program: A Study of Program Attrition. J Med Internet Res 2010;12(5):e57. [doi: 10.5185/amlett.2019.2050]
84. Forchuk C, Reiss J, Eichstedt J, Singh D, Collins K, Rudnick A, et al. The Youth-Mental Health Engagement Network: An Exploratory Pilot Study of a Smartphone and Computer-Based Personal Health Record for Youth Experiencing Depressive Symptoms. Int J Ment Health 2016;45(3):205-222. [doi: 10.1080/00207411.2016.1204823]
85. Nitsch M, Dimopoulos C, Flaschberger E, Saffran K, Kruger J, Garlock L, et al. A guided online and mobile self-help program for individuals with eating disorders: An iterative engagement and usability study. J Med Internet Res 2016;18(1):e7 [FREE Full text] [doi: 10.2196/jmir.4972]
86. Heiniger L, Smith A, Olver I, Grimison P, Klein B, Wootten A, et al. Development and pilot testing of a web‐based intervention to reduce anxiety and depression in survivors of testicular cancer. Eur J Cancer Care (Engl) 2017;26(6):1-10 [FREE Full text] [doi: 10.1111/ecc.12698]

87. Meyer N, Kerz M, Folarin A, Joyce DW, Jackson R, Karr C, et al. Capturing Rest-Activity Profiles in Schizophrenia Using Wearable and Mobile Technologies: Development, Implementation, Feasibility, and Acceptability of a Remote Monitoring Platform. JMIR Mhealth Uhealth 2018 Oct 30;6(10):e188. [doi: 10.2196/mhealth.8292]
88. Burchert S, Alkneme MS, Bird M, Carswell K, Cuijpers P, Hansen P, et al. User-Centered App Adaptation of a Low-Intensity E-Mental Health Intervention for Syrian Refugees. Front. Psychiatry 2019 Jan 25;9(663):1-18. [doi: 10.3389/fpsyt.2018.00663]
89. Görges F, Oehler C, von Hirschhausen E, Hegerl U, Rummel-Kluge C. GET.HAPPY2 - User perspectives on an internet-based self-management positive psychology intervention among persons with and without depression: Results from a retrospective survey. J Clin Psychol 2020 Jun;76(6):1030-1046. [doi: 10.1002/jclp.22886] [Medline: 31714609]
90. Kerr D, Ornelas I, Lilly M, Calhoun R, Meischke H. Participant Engagement in and Perspectives on a Web-Based Mindfulness Intervention for 9-1-1 Telecommunicators: Multimethod Study. J Med Internet Res 2019 Jun 19;21(6):e13449 [FREE Full text] [doi: 10.2196/13449] [Medline: 31219045]
91. Sayal K, Roe J, Ball H, Atha C, Kaylor-Hughes C, Guo B, et al. Feasibility of a randomised controlled trial of remotely delivered problem-solving cognitive behaviour therapy versus usual care for young people with depression and repeat self-harm: lessons learnt (e-DASH). BMC Psychiatry 2019 Jan 24;19(1):42. [doi: 10.1186/s12888-018-2005-3] [Medline: 30678674]
92. Schuster R, Kalthoff I, Walther A, Köhldorfer L, Partinger E, Berger T, et al. Effects, Adherence, and Therapists' Perceptions of Web- and Mobile-Supported Group Therapy for Depression: Mixed-Methods Study. J Med Internet Res 2019 Apr 28;21(5):e11860 [FREE Full text] [doi: 10.2196/11860] [Medline: 31066700]
93. Simblett S, Matcham F, Siddi S, Bulgari V, Barattieri di San Pietro C, Hortas López J, et al. Barriers to and Facilitators of Engagement With mHealth Technology for Remote Measurement and Management of Depression: Qualitative Analysis. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e11325. [doi: 10.2196/11325]
94. Lim MH, Rodebaugh TL, Eres R, Long KM, Penn DL, Gleeson JFM. A Pilot Digital Intervention Targeting Loneliness in Youth Mental Health. Front. Psychiatry 2019 Aug 23;10:1-13. [doi: 10.3389/fpsyt.2019.00604]
95. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012 Jun 22;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]
96. Kaczmarek LD, Kashdan TB, Kleiman EM, Baczkowski B, Enko J, Siebers A, et al. Who self-initiates gratitude interventions in daily life? An examination of intentions, curiosity, depressive symptoms, and life satisfaction. Personality and Individual Differences 2013 Oct;55(7):805-810. [doi: 10.1016/j.paid.2013.06.013]
97. Al-Asadi AM, Klein B, Meyer D. Posttreatment attrition and its predictors, attrition bias, and treatment efficacy of the anxiety online programs. J Med Internet Res 2014 Oct 14;16(10):e232 [FREE Full text] [doi: 10.2196/jmir.3513] [Medline: 25316533]
98. von Brachel R, Hötzel K, Hirschfeld G, Rieger E, Schmidt U, Kosfelder J, et al. Internet-Based Motivation Program for Women With Eating Disorders: Eating Disorder Pathology and Depressive Mood Predict Dropout. J Med Internet Res 2014 Mar 31;16(3):e92. [doi: 10.2196/jmir.3104]
99. Farrer LM, Griffiths KM, Christensen H, Mackinnon AJ, Batterham PJ. Predictors of Adherence and Outcome in Internet-Based Cognitive Behavior Therapy Delivered in a Telephone Counseling Setting. Cogn Ther Res 2013 Oct 27;38(3):358-367. [doi: 10.1007/s10608-013-9589-1]
100. Habibović M, Cuijpers P, Alings M, van DVP, Theuns D, Bouwels L, et al. Attrition and adherence in a WEB-Based Distress Management Program for Implantable Cardioverter defibrillator Patients (WEBCARE): randomized controlled trial. J Med Internet Res 2014;16(2):e52 [FREE Full text] [doi: 10.2196/jmir.2809] [Medline: 24583632]
101. Grubbs KM, Fortney JC, Pyne JM, Hudson T, Moore WM, Custer P, et al. Predictors of Initiation and Engagement of Cognitive Processing Therapy Among Veterans With PTSD Enrolled in Collaborative Care. Journal of Traumatic Stress 2015 Dec 01;28(6):580-584. [doi: 10.1002/jts.22049]
102. Gould C, Zapata A, Bruce J, Bereknyei Merrell S, Wetherell J, O'Hara R, et al. Development of a video-delivered relaxation treatment of late-life anxiety for veterans. Int. Psychogeriatr 2017 Jun 08;29(10):1633-1645. [doi: 10.1017/s1041610217000928]
103. Feijt MA, de Kort YAW, Bongers IM, IJsselsteijn WA. Perceived Drivers and Barriers to the Adoption of eMental Health by Psychologists: The Construction of the Levels of Adoption of eMental Health Model. J Med Internet Res 2018 Apr 24;20(4):e153 [FREE Full text] [doi: 10.2196/jmir.9485] [Medline: 29691215]
104. Berry N, Lobban F, Bucci S. A qualitative exploration of service user views about using digital health interventions for self-management in severe mental health problems. BMC Psychiatry 2019 Jan 21;19(1):35 [FREE Full text] [doi: 10.1186/s12888-018-1979-1] [Medline: 30665384]
105. Deen TL, Fortney JC, Schroeder G. Patient acceptance of and initiation and engagement in telepsychotherapy in primary care. Psychiatr Serv 2013 Apr 1;64(4):380-384 [FREE Full text] [doi: 10.1176/appi.ps.201200198] [Medline: 23370530]
106. Fonseca A, Gorayeb R, Canavarro MC. Women's use of online resources and acceptance of e-mental health tools during the perinatal period. International Journal of Medical Informatics 2016 Oct;94:228-236. [doi: 10.1016/j.ijmedinf.2016.07.016]

107. O'Dea B, Achilles M, Werner-Seidler A, Batterham P, Calear A, Perry Y, et al. Adolescents' Perspectives on a Mobile App for Relationships: Cross-Sectional Survey. JMIR Mhealth Uhealth 2018 Mar 08;6(3):e56 [FREE Full text] [doi: 10.2196/mhealth.8831] [Medline: 29519775]
108. Adkins EC, Zalta AK, Boley RA, Glover A, Karnik NS, Schueller SM. Exploring the potential of technology-based mental health services for homeless youth: A qualitative study. Psychol Serv 2017 Dec;14(2):238-245 [FREE Full text] [doi: 10.1037/ser0000120] [Medline: 28481610]
109. Apolinário-Hagen J, Hennemann S, Fritsche L, Drüge M, Breil B. Determinant Factors of Public Acceptance of Stress Management Apps: Survey Study. JMIR Ment Health 2019 Nov 07;6(11):e15373 [FREE Full text] [doi: 10.2196/15373] [Medline: 31697243]
110. Nitsch M, Dimopoulos CN, Flaschberger E, Saffran K, Kruger JF, Garlock L, et al. A Guided Online and Mobile Self-Help Program for Individuals With Eating Disorders: An Iterative Engagement and Usability Study. J Med Internet Res 2016 Jan 11;18(1):e7 [FREE Full text] [doi: 10.2196/jmir.4972] [Medline: 26753539]
111. East M, Havard B, Hastings N. Mental Health Mobile Apps' Instruction: Technology Adoption Theories Applied in a Mixed Methods Study of Counseling Faculty. Journal of Technology in Human Services 2016 Oct 21;34(4):301-325. [doi: 10.1080/15228835.2016.1233842]
112. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital Mental Health Interventions for Depression, Anxiety, and Enhancement of Psychological Well-Being Among College Students: Systematic Review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
113. Huerta-Ramos E, Marcó-García S, Escobar-Villegas MS, Rubio-Abadal E, Ochoa S, Grasa Bello EM, et al. m-RESIST, a complete m-Health solution for patients with treatmentresistant schizophrenia: a qualitative study of user needs and acceptability in the Barcelona metropolitan area. Actas Esp Psiquiatr 2017 Nov;45(6):277-289 [FREE Full text] [Medline: 29199762]
114. Switsers L, Dauwe A, Vanhoudt A, Van Dyck H, Lombaerts K, Oldenburg J. Users' Perspectives on mHealth Self-Management of Bipolar Disorder: Qualitative Focus Group Study. JMIR Mhealth Uhealth 2018 May 02;6(5):e108. [doi: 10.2196/mhealth.9529]
115. Diez-Canseco F, Toyama M, Ipince A, Perez-Leon S, Cavero V, Araya R, et al. Integration of a Technology-Based Mental Health Screening Program Into Routine Practices of Primary Health Care Services in Peru (The Allillanchu Project): Development and Implementation. J Med Internet Res 2018 Mar 15;20(3):e100 [FREE Full text] [doi: 10.2196/jmir.9208] [Medline: 29588272]
116. Gindidis S, Stewart S, Roodenburg J. Psychologists' motivations for integrating apps into therapy with secondary school-aged young people. J. Psychol. Couns. Sch 2019 Oct 21;30(1):2-12. [doi: 10.1017/jgc.2019.22]
117. Allan S, Bradstreet S, Mcleod H, Farhall J, Lambrou M, Gleeson J, et al. Developing a Hypothetical Implementation Framework of Expectations for Monitoring Early Signs of Psychosis Relapse Using a Mobile App: Qualitative Study. J Med Internet Res 2019 Oct 24;21(10):e14366. [doi: 10.2196/14366]
118. Newman L, Bidargaddi N, Schrader G. Service providers' experiences of using a telehealth network 12 months after digitisation of a large Australian rural mental health service. International Journal of Medical Informatics 2016 Oct;94:8-20. [doi: 10.1016/j.ijmedinf.2016.05.006]
119. Thorsen MM, Patena JV, Guthrie KM, Spirito A, Ranney ML. Using High-Risk Adolescents′ Voices to Develop a Comprehensible Cognitive Behavioral Therapy–Based Text-Message Program. Behavioral Medicine 2016 Nov 21;44(2):89-99. [doi: 10.1080/08964289.2016.1223597]
120. Chęć M, Ligocka M, Maciejewska M, Samochowiec J, Łodygowska E, Samochowiec A. On-line psychological support in the evaluation of specialists and future specialists in Poland. Computers in Human Behavior 2016 Nov;64:703-709. [doi: 10.1016/j.chb.2016.07.015]
121. Jordan SE, Shearer EM. An exploration of supervision delivered via clinical video telehealth (CVT). Training and Education in Professional Psychology 2019 Nov;13(4):323-330. [doi: 10.1037/tep0000245]
122. Bucci S, Berry N, Morris R, Berry K, Haddock G, Lewis S, et al. "They Are Not Hard-to-Reach Clients. We Have Just Got Hard-to-Reach Services." Staff Views of Digital Health Tools in Specialist Mental Health Services. Front Psychiatry 2019;10:344 [FREE Full text] [doi: 10.3389/fpsyt.2019.00344] [Medline: 31133906]
123. Moessner M, Minarik C, Özer F, Bauer S. Can an internet-based program for the prevention and early intervention in eating disorders facilitate access to conventional professional healthcare? J Ment Health 2016 Oct;25(5):441-447. [doi: 10.3109/09638237.2016.1139064] [Medline: 26850624]
124. Meglic M, Furlan M, Kuzmanic M, Kozel D, Baraga D, Kuhar I, et al. Feasibility of an eHealth Service to Support Collaborative Depression Care: Results of a Pilot Study. J Med Internet Res 2010 Dec 19;12(5):e63. [doi: 10.2196/jmir.1510]
125. Forchuk C, Reiss J, O'Regan T, Ethridge P, Donelle L, Rudnick A. Client perceptions of the mental health engagement network: a qualitative analysis of an electronic personal health record. BMC Psychiatry 2015 Oct 14;15:250 [FREE Full text] [doi: 10.1186/s12888-015-0614-7] [Medline: 26467210]
126. Wozney L, Newton AS, Gehring ND, Bennett K, Huguet A, Hartling L, et al. Implementation of eMental Health care: viewpoints from key informants from organizations and agencies with eHealth mandates. BMC Med Inform Decis Mak 2017 Jun 2;17(1):1-15. [doi: 10.1186/s12911-017-0474-9]

127. Wallin EEK, Mattsson S, Olsson EMG. The Preference for Internet-Based Psychological Interventions by Individuals Without Past or Current Use of Mental Health Treatment Delivered Online: A Survey Study With Mixed-Methods Analysis. JMIR Ment Health 2016;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]
128. Lal S, Dell'Elce J, Tucci N, Fuhrer R, Tamblyn R, Malla A. Preferences of Young Adults With First-Episode Psychosis for Receiving Specialized Mental Health Services Using Technology: A Survey Study. JMIR Ment Health 2015;2(2):e18 [FREE Full text] [doi: 10.2196/mental.4400] [Medline: 26543922]
129. Woolderink M, Bindels JA, Evers SM, Paulus AT, van Asselt AD, van Schayck OC. An Online Health Prevention Intervention for Youth with Addicted or Mentally Ill Parents: Experiences and Perspectives of Participants and Providers from a Randomized Controlled Trial. J Med Internet Res 2015 Dec 02;17(12):e274. [doi: 10.2196/jmir.4817]
130. Laurie J, Blandford A. Making time for mindfulness. International Journal of Medical Informatics 2016 Dec;96:38-50. [doi: 10.1016/j.ijmedinf.2016.02.010]
131. Mitchell AE, Whittingham K, Steindl S, Kirby J. Feasibility and acceptability of a brief online self-compassion intervention for mothers of infants. Arch Womens Ment Health 2018 Oct;21(5):553-561. [doi: 10.1007/s00737-018-0829-y] [Medline: 29550890]
132. Bengtsson J, Nordin S, Carlbring P. Therapists' Experiences of Conducting Cognitive Behavioural Therapy Online vis-à-vis Face-to-Face. Cognitive Behaviour Therapy 2015 Jun 19;44(6):470-479. [doi: 10.1080/16506073.2015.1053408]
133. Walsh S, Kaselionyte J, Taylor S, Priebe S. What might affect acceptability of online positive psychology interventions for depression: a qualitative study on patient expectations'. BMC Psychiatry 2018 Jul 27;18(1):240 [FREE Full text] [doi: 10.1186/s12888-018-1812-x] [Medline: 30053854]
134. Urech A, Krieger T, Möseneder L, Biaggi A, Vincent A, Poppe C, et al. A patient perspective on advantages and disadvantages of blended cognitive behaviour therapy for depression: A qualitative content analysis. Psychotherapy Research 2018 Jan 31;29(8):986-998. [doi: 10.1080/10503307.2018.1430910]
135. Anastasiadou D, Folkvord F, Serrano-Troncoso E, Lupiañez-Villanueva F. Mobile Health Adoption in Mental Health: User Experience of a Mobile Health App for Patients With an Eating Disorder. JMIR Mhealth Uhealth 2019 May 31;7(6):e12920. [doi: 10.2196/12920]
136. Saberi P, Dawson Rose C, Wootton A, Ming K, Legnitto D, Jeske M, et al. Use of technology for delivery of mental health and substance use services to youth living with HIV: a mixed-methods perspective. AIDS Care 2020 Aug;32(8):931-939. [doi: 10.1080/09540121.2019.1622637] [Medline: 31132864]
137. Carolan S, de Visser RO. Employees' Perspectives on the Facilitators and Barriers to Engaging With Digital Mental Health Interventions in the Workplace: Qualitative Study. JMIR Ment Health 2018 Jan 19;5(1):e8. [doi: 10.2196/mental.9146]
138. Ashwick R, Turgoose D, Murphy D. Exploring the acceptability of delivering Cognitive Processing Therapy (CPT) to UK veterans with PTSD over Skype: a qualitative study. European Journal of Psychotraumatology 2019 Feb 04;10(1):1573128. [doi: 10.1080/20008198.2019.1573128]
139. Jonathan G, Carpenter-Song EA, Brian RM, Ben-Zeev D. Life with FOCUS: A qualitative evaluation of the impact of a smartphone intervention on people with serious mental illness. Psychiatr Rehabil J 2019 Jun;42(2):182-189. [doi: 10.1037/prj0000337] [Medline: 30589278]
140. Lal S, Nguyen V, Theriault J. Seeking mental health information and support online: experiences and perspectives of young people receiving treatment for first-episode psychosis. Early Intervention in Psychiatry 2016 Jan 26;12(3):324-330. [doi: 10.1111/eip.12317] [Medline: 26810026]
141. Pretorius C, Chambers D, Cowan B, Coyle D. Young People Seeking Help Online for Mental Health: Cross-Sectional Survey Study. JMIR Ment Health 2019 Aug 26;6(8):e13524 [FREE Full text] [doi: 10.2196/13524] [Medline: 31452519]
142. Ranney M, Thorsen M, Patena JV, Cunningham R, Boyer E, Walton M, et al. "You need to get them where they feel it": Conflicting perspectives on how to maximize the structure of text-message psychological interventions for adolescents. : IEEE Computer Society; 2015 Presented at: 48th Hawaii Int Conf Syst Sci; Jan 5-8; Kauai, Hawaii p. 3247-3255. [doi: 10.1109/hicss.2015.391]
143. Casey LM, Joy A, Clough BA. The impact of information on attitudes toward e-mental health services. Cyberpsychol Behav Soc Netw 2013 Aug;16(8):593-598. [doi: 10.1089/cyber.2012.0515] [Medline: 23679567]
144. Sinha Deb K, Tuli A, Sood M, Chadda R, Verma R, Kumar S, et al. Is India ready for mental health apps (MHApps)? A quantitative-qualitative exploration of caregivers' perspective on smartphone-based solutions for managing severe mental illnesses in low resource settings. PLoS One 2018;13(9):e0203353 [FREE Full text] [doi: 10.1371/journal.pone.0203353] [Medline: 30231056]
145. Chadi N, Weisbaum E, Malboeuf-Hurtubise C, Kohut S, Viner C, Kaufman M, et al. Can the Mindful Awareness and Resilience Skills for Adolescents (MARS-A) Program Be Provided Online? Voices from the Youth. Children (Basel) 2018;5(9):115. [doi: 10.20944/preprints201807.0545.v1]
146. Kubo A, Altschuler A, Kurtovich E, Hendlish S, Laurent CA, Kolevska T, et al. A Pilot Mobile-Based Mindfulness Intervention for Cancer Patients and Their Informal Caregivers. Mindfulness 2018 Mar 24;9(6):1885-1894. [doi: 10.1007/s12671-018-0931-2]

147. Postel M, de Haan HA, ter Huurne ED, Becker E, de Jong CAJ. Effectiveness of a web-based intervention for problem drinkers and reasons for dropout: randomized controlled trial. J Med Internet Res 2010 Dec 16;12(4):e68 [FREE Full text] [doi: 10.2196/jmir.1642] [Medline: 21163776]
148. Ashford MT, Olander EK, Rowe H, Fisher JR, Ayers S. Feasibility and Acceptability of a Web-Based Treatment with Telephone Support for Postpartum Women With Anxiety: Randomized Controlled Trial. JMIR Ment Health 2018 Apr 20;5(2):e19. [doi: 10.2196/mental.9106]
149. Chen AT, Wu S, Tomasino KN, Lattie EG, Mohr DC. A multi-faceted approach to characterizing user behavior and experience in a digital mental health intervention. Journal of Biomedical Informatics 2019 Jun;94:103187. [doi: 10.1016/j.jbi.2019.103187]
150. Wallin E, Norlund F, Olsson EMG, Burell G, Held C, Carlsson T. Treatment Activity, User Satisfaction, and Experienced Usability of Internet-Based Cognitive Behavioral Therapy for Adults With Depression and Anxiety After a Myocardial Infarction: Mixed-Methods Study. J Med Internet Res 2018 Mar 16;20(3):e87 [FREE Full text] [doi: 10.2196/jmir.9690] [Medline: 29549067]
151. Schuster R, Fichtenbauer I, Sparr VM, Berger T, Laireiter A. Feasibility of a blended group treatment (bGT) for major depression: uncontrolled interventional study in a university setting. BMJ Open 2018 Dec 12;8(3):e018412 [FREE Full text] [doi: 10.1136/bmjopen-2017-018412] [Medline: 29530905]
152. Chan C, West S, Glozier N. Commencing and Persisting With a Web-Based Cognitive Behavioral Intervention for Insomnia: A Qualitative Study of Treatment Completers. J Med Internet Res 2017 Feb 10;19(2):e37 [FREE Full text] [doi: 10.2196/jmir.5639] [Medline: 28188124]
153. Dodd A, Mallinson S, Griffiths M, Morriss R, Jones S, Lobban F. Users' experiences of an online intervention for bipolar disorder: important lessons for design and evaluation. Evid Based Ment Health 2017 Nov;20(4):133-139 [FREE Full text] [doi: 10.1136/eb-2017-102754] [Medline: 29061886]
154. Powell L, Parker J, Harpin V. ADHD: Is There an App for That? A Suitability Assessment of Apps for the Parents of Children and Young People With ADHD. JMIR Mhealth Uhealth 2017 Oct 13;5(10):e149. [doi: 10.2196/mhealth.7941]
155. Lundgren J, Johansson P, Jaarsma T, Andersson G, Kärner Köhler A. Patient Experiences of Web-Based Cognitive Behavioral Therapy for Heart Failure and Depression: Qualitative Study. J Med Internet Res 2018 Sep 05;20(9):e10302 [FREE Full text] [doi: 10.2196/10302] [Medline: 30185405]
156. Huis in het Veld JG, Verkaik R, van Meijel B, Verkade P, Werkman W, Hertogh CM, et al. Self-Management Support and eHealth When Managing Changes in Behavior and Mood of a Relative With Dementia: An Asynchronous Online Focus Group Study of Family Caregivers' Needs. Research in Gerontological Nursing 2018 Mar 02;11(3):151-159. [doi: 10.3928/19404921-20180216-01]
157. Hamblen J, Grubaugh A, Davidson T, Borkman A, Bunnell B, Ruggiero K. An Online Peer Educational Campaign to Reduce Stigma and Improve Help Seeking in Veterans with Posttraumatic Stress Disorder. Telemed J E Health 2019 Jan;25(1):41-47. [doi: 10.1089/tmj.2017.0305] [Medline: 29746232]
158. Schuster R, Sigl S, Berger T, Laireiter A. Patients' Experiences of Web- and Mobile-Assisted Group Therapy for Depression and Implications of the Group Setting: Qualitative Follow-Up Study. JMIR Ment Health 2018 Jul 11;5(3):e49 [FREE Full text] [doi: 10.2196/mental.9613] [Medline: 29997106]
159. Similä H, Immonen M, Toska-Tervola J, Enwald H, Keränen N, Kangas M, et al. Feasibility of mobile mental wellness training for older adults. Geriatr Nurs 2018;39(5):499-505. [doi: 10.1016/j.gerinurse.2018.02.001] [Medline: 29530292]
160. Wachtler C, Coe A, Davidson S, Fletcher S, Mendoza A, Sterling L, et al. Development of a Mobile Clinical Prediction Tool to Estimate Future Depression Severity and Guide Treatment in Primary Care: User-Centered Design. JMIR Mhealth Uhealth 2018 Apr 23;6(4):e95 [FREE Full text] [doi: 10.2196/mhealth.9502] [Medline: 29685864]
161. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of Mobile Apps for Depression and Anxiety: Cross-Sectional Survey Study on Patient Interest and Barriers to Engagement. JMIR Ment Health 2019 Jan 25;6(1):e11334 [FREE Full text] [doi: 10.2196/11334] [Medline: 30681968]
162. Hartmann R, Sander C, Lorenz N, Böttger D, Hegerl U. Utilization of Patient-Generated Data Collected Through Mobile Devices: Insights From a Survey on Attitudes Toward Mobile Self-Monitoring and Self-Management Apps for Depression. JMIR Ment Health 2019 Apr 03;6(4):e11671 [FREE Full text] [doi: 10.2196/11671] [Medline: 30942693]
163. Kenny R, Dooley B, Fitzgerald A. Feasibility of "CopeSmart": A Telemental Health App for Adolescents. JMIR Ment Health 2015 Aug;2(3):e22 [FREE Full text] [doi: 10.2196/mental.4370] [Medline: 26552425]
164. Anderson AP, Fellows AM, Binsted KA, Hegel MT, Buckey JC. Autonomous, Computer-Based Behavioral Health Countermeasure Evaluation at HI-SEAS Mars Analog. Aerospace Medicine and Human Performance 2016 Nov 01;87(11):912-920. [doi: 10.3357/amhp.4676.2016]
165. Stiles-Shields C, Montague E, Lattie EG, Kwasny MJ, Mohr DC. What might get in the way: Barriers to the use of apps for depression. DIGITAL HEALTH 2017 Jun 08;3:205520761771382. [doi: 10.1177/2055207617713827]
166. Gonsalves P, Hodgson E, Kumar A, Aurora T, Chandak Y, Sharma R, et al. Design and Development of the "POD Adventures" Smartphone Game: A Blended Problem-Solving Intervention for Adolescent Mental Health in India. Front Public Health 2019;7:238 [FREE Full text] [doi: 10.3389/fpubh.2019.00238] [Medline: 31508404]

167. Povey J, Mills PPJR, Dingwall KM, Lowell A, Singer J, Rotumah D, et al. Acceptability of Mental Health Apps for Aboriginal and Torres Strait Islander Australians: A Qualitative Study. J Med Internet Res 2016;18(3):e65 [FREE Full text] [doi: 10.2196/jmir.5314] [Medline: 26969043]
168. Hunter-Jones JJ, Gilliam SM, Carswell AL, Hansen NB. Assessing the Acceptability of a Mindfulness-Based Cognitive Therapy Intervention for African-American Women Living with HIV/AIDS. J. Racial and Ethnic Health Disparities 2019 Jul 22;6(6):1157-1166. [doi: 10.1007/s40615-019-00617-5]
169. Henshall C, Marzano L, Smith K, Attenburrow M, Puntis S, Zlodre J, et al. A web-based clinical decision tool to support treatment decision-making in psychiatry: a pilot focus group study with clinicians, patients and carers. BMC Psychiatry 2017 Jul 21;17(1):265. [doi: 10.1186/s12888-017-1406-z] [Medline: 28732477]
170. Murnane EL, Cosley D, Chang P, Guha S, Frank E, Gay G, et al. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health. J Am Med Inform Assoc 2016 May;23(3):477-484. [doi: 10.1093/jamia/ocv165] [Medline: 26911822]
171. Lord S, Moore SK, Ramsey A, Dinauer S, Johnson K. Implementation of a Substance Use Recovery Support Mobile Phone App in Community Settings: Qualitative Study of Clinician and Staff Perspectives of Facilitators and Barriers. JMIR Ment Health 2016 Jun 28;3(2):e24 [FREE Full text] [doi: 10.2196/mental.4927] [Medline: 27352884]
172. Caplan S, Sosa Lovera A, Reyna Liberato P. A feasibility study of a mental health mobile app in the Dominican Republic: the untold story. Int J Mental Health 2019 Apr 04;47(4):311-345. [doi: 10.1080/00207411.2018.1553486]
173. Apolinário-Hagen J, Vehreschild V, Alkoudmani RM. Current Views and Perspectives on E-Mental Health: An Exploratory Survey Study for Understanding Public Attitudes Toward Internet-Based Psychotherapy in Germany. JMIR Ment Health 2017 Feb 23;4(1):e8. [doi: 10.2196/mental.6375]
174. Kumarasiri J, Jubb C. Framing of Climate Change Impacts and Use of Management Accounting Practices. AAMJAF 2018 Feb 28;13(2):45-68. [doi: 10.21315/aamjaf2017.13.2.3]
175. Huerta-Ramos E, Escobar-Villegas MS, Rubinstein K, Unoka ZS, Grasa E, Hospedales M, M-RESIST Group, et al. Measuring Users' Receptivity Toward an Integral Intervention Model Based on mHealth Solutions for Patients With Treatment-Resistant Schizophrenia (m-RESIST): A Qualitative Study. JMIR Mhealth Uhealth 2016 Sep 28;4(3):e112 [FREE Full text] [doi: 10.2196/mhealth.5716] [Medline: 27682896]
176. Giroux D, Bacon S, King DK, Dulin P, Gonzalez V. Examining Perceptions of a Smartphone-Based Intervention System for Alcohol Use Disorders. Telemedicine and e-Health 2014 Oct;20(10):923-929. [doi: 10.1089/tmj.2013.0222]
177. Gonzales R, Douglas Anglin M, Glik D. Exploring the feasibility of text messaging to support substance abuse recovery among youth in treatment. Health Educ Res 2014 Feb;29(1):13-22 [FREE Full text] [doi: 10.1093/her/cyt094] [Medline: 24038196]
178. Lattie EG, Ho J, Sargent E, Tomasino KN, Smith JD, Brown CH, et al. Teens Engaged in Collaborative Health: The Feasibility and Acceptability of an Online Skill-Building Intervention for Adolescents at Risk for Depression. Internet Interv 2017 Jun;8:15-26 [FREE Full text] [doi: 10.1016/j.invent.2017.02.003] [Medline: 28584734]
179. Miatello A, Mulvale G, Hackett C, Mulvale A, Kutty A, Alshazly F. Data Elicited Through Apps for Health Systems Improvement: Lessons From Using the myEXP Suite of Smartphone and Web Apps. International Journal of Qualitative Methods 2018 Oct;17(1):1-13. [doi: 10.1177/1609406918798433]
180. O'Brien KHM, Battalen A, Sellers C, Spirito A, Yen S, Maneta E, et al. An mHealth approach to extend a brief intervention for adolescent alcohol use and suicidal behavior: Qualitative analyses of adolescent and parent feedback. J Technol Hum Serv 2019;37(4):255-285 [FREE Full text] [doi: 10.1080/15228835.2018.1561347] [Medline: 31814806]
181. Sreejith G, Menon V. Mobile Phones as a Medium of Mental Health Care Service Delivery: Perspectives and Barriers among Patients with Severe Mental Illness. Indian Journal of Psychological Medicine 2019 Oct 01;41(5):428-433. [doi: 10.4103/ijpsym.ijpsym_333_18]
182. Burda C, Haack M, Duarte A, Alemi F. Medication adherence among homeless patients: a pilot study of cell phone effectiveness. J Am Acad Nurse Pract 2012 Nov;24(11):675-681. [doi: 10.1111/j.1745-7599.2012.00756.x] [Medline: 23088699]
183. Berry N, Bucci S, Lobban F. Use of the Internet and Mobile Phones for Self-Management of Severe Mental Health Problems: Qualitative Study of Staff Views. JMIR Ment Health 2017 Nov 01;4(4):e52. [doi: 10.2196/mental.8311] [Medline: 29092809]
184. Whealin JM, Jenchura EC, Wong AC, Zulman DM. How Veterans With Post-Traumatic Stress Disorder and Comorbid Health Conditions Utilize eHealth to Manage Their Health Care Needs: A Mixed-Methods Analysis. J Med Internet Res 2016 Oct 26;18(10):e280 [FREE Full text] [doi: 10.2196/jmir.5594] [Medline: 27784650]
185. Nicholas J, Fogarty AS, Boydell K, Christensen H. The Reviews Are in: A Qualitative Content Analysis of Consumer Perspectives on Apps for Bipolar Disorder. J Med Internet Res 2017 Apr 07;19(4):e105 [FREE Full text] [doi: 10.2196/jmir.7273] [Medline: 28389420]
186. Fortuna KL, Storm M, Naslund JA, Chow P, Aschbrenner KA, Lohman MC, et al. Certified Peer Specialists and Older Adults With Serious Mental Illness Perspectives of the Impact of a Peer-Delivered and Technology-Supported Self-Management Intervention. The Journal of Nervous and Mental Disease 2018;206(11):875-881. [doi: 10.1097/nmd.0000000000000896]

187. Pung A, Fletcher SL, Gunn JM. Mobile App Use by Primary Care Patients to Manage Their Depressive Symptoms: Qualitative Study. J Med Internet Res 2018 Sep 27;20(9):e10035 [FREE Full text] [doi: 10.2196/10035] [Medline: 30262449]
188. Carr A, Jones J, Mikulich Gilbertson S, Laudenslager M, Kutner J, Kilbourn K, et al. Impact of a Mobilized Stress Management Program (Pep-Pal) for Caregivers of Oncology Patients: Mixed-Methods Study. JMIR Cancer 2019 May 03;5(1):e11406 [FREE Full text] [doi: 10.2196/11406] [Medline: 31066678]
189. Jarvis MA, Chipps J, Padmanabhanunni A. "This phone saved my life": Older persons' experiences and appraisals of an mHealth intervention aimed at addressing loneliness. Journal of Psychology in Africa 2019 May 05;29(2):159-166. [doi: 10.1080/14330237.2019.1594650]
190. Klein P, Lawn S, Tsourtos G, van Agteren J. Tailoring of a Smartphone Smoking Cessation App (Kick.it) for Serious Mental Illness Populations: Qualitative Study. JMIR Hum Factors 2019 Sep 03;6(3):e14023 [FREE Full text] [doi: 10.2196/14023] [Medline: 31482850]
191. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and Facilitators of Engagement With Remote Measurement Technology for Managing Health: Systematic Review and Content Analysis of Findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
192. Aardoom J, Dingemans A, van Ginkel JR, Spinhoven P, Van Furth EF, Van den Akker-van Marle ME. Cost-utility of an internet-based intervention with or without therapist support in comparison with a waiting list for individuals with eating disorder symptoms: a randomized controlled trial. Int J Eat Disord 2016 Dec;49(12):1068-1076. [doi: 10.1002/eat.22587] [Medline: 27441418]
193. Carolan S, Harris P, Greenwood K, Cavanagh K. Increasing engagement with an occupational digital stress management program through the use of an online facilitated discussion group: Results of a pilot randomised controlled trial. Internet Interv 2017 Dec;10:1-11 [FREE Full text] [doi: 10.1016/j.invent.2017.08.001] [Medline: 30135747]
194. Ho J, Corden ME, Caccamo L, Tomasino KN, Duffecy J, Begale M, et al. Design and evaluation of a peer network to support adherence to a web-based intervention for adolescents. Internet Interventions 2016 Nov;6:50-56. [doi: 10.1016/j.invent.2016.09.005]
195. Espinosa H, Carrasco A, Moessner M, Cáceres C, Gloger S, Rojas G, et al. Acceptability Study of "Ascenso": An Online Program for Monitoring and Supporting Patients with Depression in Chile. Telemed J E Health 2016 Jul;22(7):577-583. [doi: 10.1089/tmj.2015.0124] [Medline: 26741190]
196. Clarke J, Proudfoot J, Whitton A, Birch M, Boyd M, Parker G, et al. Therapeutic Alliance With a Fully Automated Mobile Phone and Web-Based Intervention: Secondary Analysis of a Randomized Controlled Trial. JMIR Mental Health 2016 Feb 25;3(1):e10. [doi: 10.2196/mental.4656]
197. Muuraiskangas S, Harjumaa M, Kaipainen K, Ermes M. Process and Effects Evaluation of a Digital Mental Health Intervention Targeted at Improving Occupational Well-Being: Lessons From an Intervention Study With Failed Adoption. JMIR Ment Health 2016 May 11;3(2):e13 [FREE Full text] [doi: 10.2196/mental.4465] [Medline: 27170553]
198. Schroeder J, Wilks C, Rowan K, Toledo A, Paradiso A, Czerwinski M, et al. Pocket skills: A conversational Mobile web app to support dialectical behavioral therapy. : Association for Computing Machinery; 2018 Presented at: CHI'18; April 21-26; Montreal, Canada p. 1-15. [doi: 10.1145/3173574.3173972]
199. Kim E, Gellis ZD, Bradway C, Kenaley B. Key Determinants to using Telehealth Technology to Serve Medically Ill and Depressed Homebound Older Adults. Journal of Gerontological Social Work 2018 Jul 24;62(4):451-474. [doi: 10.1080/01634372.2018.1499575]
200. Terp M, Jørgensen R, Laursen B, Mainz J, Bjørnes CD. A Smartphone App to Foster Power in the Everyday Management of Living With Schizophrenia: Qualitative Analysis of Young Adults' Perspectives. JMIR Ment Health 2018 Oct 01;5(4):e10157 [FREE Full text] [doi: 10.2196/10157] [Medline: 30274966]
201. Williams A, Fossey E, Farhall J, Foley F, Thomas N. Recovery After Psychosis: Qualitative Study of Service User Experiences of Lived Experience Videos on a Recovery-Oriented Website. JMIR Ment Health 2018 May 08;5(2):e37 [FREE Full text] [doi: 10.2196/mental.9934] [Medline: 29739737]
202. Saunders K, Bilderbeck A, Panchal P, Atkinson L, Geddes J, Goodwin G. Experiences of remote mood and activity monitoring in bipolar disorder: A qualitative study. Eur Psychiatry 2017 Mar;41:115-121 [FREE Full text] [doi: 10.1016/j.eurpsy.2016.11.005] [Medline: 28135594]
203. Shrier LA, Spalding A. "Just Take a Moment and Breathe and Think": Young Women with Depression Talk about the Development of an Ecological Momentary Intervention to Reduce Their Sexual Risk. Journal of Pediatric and Adolescent Gynecology 2017 Feb;30(1):116-122. [doi: 10.1016/j.jpag.2016.08.009]
204. Cernvall M, Sveen J, Bergh Johannesson K, Arnberg F. A pilot study of user satisfaction and perceived helpfulness of the Swedish version of the mobile app PTSD Coach. Eur J Psychotraumatol 2018;9(Suppl 1):1472990 [FREE Full text] [doi: 10.1080/20008198.2018.1472990] [Medline: 29805783]
205. Edbrooke-Childs J, Edridge C, Averill P, Delane L, Hollis C, Craven MP, et al. A Feasibility Trial of Power Up: Smartphone App to Support Patient Activation and Shared Decision Making for Mental Health in Young People. JMIR Mhealth Uhealth 2019 Jun 04;7(6):e11677. [doi: 10.2196/11677]
206. Schlosser D, Campellone T, Kim D, Truong B, Vergani S, Ward C, et al. Feasibility of PRIME: A Cognitive Neuroscience-Informed Mobile App Intervention to Enhance Motivated Behavior and Improve Quality of Life in Recent

Onset Schizophrenia. JMIR Res Protoc 2016 Apr 28;5(2):e77 [FREE Full text] [doi: 10.2196/resprot.5450] [Medline: 27125771]
207. Garrido S, Cheers D, Boydell K, Nguyen QV, Schubert E, Dunne L, et al. Young People's Response to Six Smartphone Apps for Anxiety and Depression: Focus Group Study. JMIR Ment Health 2019 Oct 02;6(10):e14385 [FREE Full text] [doi: 10.2196/14385] [Medline: 31579023]
208. Boggs J, Beck A, Felder J, Dimidjian S, Metcalf C, Segal ZV. Web-based intervention in mindfulness meditation for reducing residual depressive symptoms and relapse prophylaxis: a qualitative study. J Med Internet Res 2014 Mar 24;16(3):e87 [FREE Full text] [doi: 10.2196/jmir.3129] [Medline: 24662625]
209. Pierce B, Twohig MP, Levin ME. Perspectives on the use of acceptance and commitment therapy related mobile apps: Results from a survey of students and professionals. Journal of Contextual Behavioral Science 2016 Oct;5(4):215-224. [doi: 10.1016/j.jcbs.2016.08.001]
210. Puszka S, Dingwall KM, Sweet M, Nagel T. E-Mental Health Innovations for Aboriginal and Torres Strait Islander Australians: A Qualitative Study of Implementation Needs in Health Services. JMIR Ment Health 2016 Sep 19;3(3):e43 [FREE Full text] [doi: 10.2196/mental.5837] [Medline: 27644259]
211. Bennett-Levy J, Singer J, DuBois S, Hyde K. Translating E-Mental Health Into Practice: What Are the Barriers and Enablers to E-Mental Health Implementation by Aboriginal and Torres Strait Islander Health Professionals? J Med Internet Res 2017 Jan 11;19(1):e1 [FREE Full text] [doi: 10.2196/jmir.6269] [Medline: 28077347]
212. Ye J, Shim R, Lukaszewski T, Yun K, Kim SH, Rust G. Telepsychiatry Services for Korean Immigrants. Telemedicine and e-Health 2012 Dec;18(10):797-802. [doi: 10.1089/tmj.2012.0041]
213. Koppe H, van de Mortel TF, Ahern CM. How effective and acceptable is Web 2.0 Balint group participation for general practitioners and general practitioner registrars in regional Australia? A pilot study. Aust J Rural Health 2016 Feb;24(1):16-22 [FREE Full text] [doi: 10.1111/ajr.12212] [Medline: 26114400]
214. Bohleber L, Crameri A, Eich-Stierli B, Telesko R, von WA. Can We Foster a Culture of Peer Support and Promote Mental Health in Adolescence Using a Web-Based App? A Control Group Study. JMIR Ment Health 2016 Sep 23;3(3):e45 [FREE Full text] [doi: 10.2196/mental.5597] [Medline: 27663691]
215. Naslund JA, Aschbrenner KA, Bartels SJ. Wearable devices and smartphones for activity tracking among people with serious mental illness. Mental Health and Physical Activity 2016 Mar;10:10-17. [doi: 10.1016/j.mhpa.2016.02.001]
216. Rickard N, Arjmand H, Bakker D, Seabrook E. Development of a Mobile Phone App to Support Self-Monitoring of Emotional Well-Being: A Mental Health Digital Innovation. JMIR Ment Health 2016 Nov 23;3(4):e49 [FREE Full text] [doi: 10.2196/mental.6202] [Medline: 27881358]
217. Reger GM, Browne KC, Campellone TR, Simons C, Kuhn E, Fortney JC, et al. Barriers and facilitators to mobile application use during PTSD treatment: Clinician adoption of PE coach. Professional Psychology: Research and Practice 2017 Dec;48(6):510-517. [doi: 10.1037/pro0000153] [Medline: 28425725]
218. Whealin JM, King L, Shore P, Spira J. Diverse veterans' pre- and post-intervention perceptions of home telemental health for posttraumatic stress disorder delivered via tablet. Int J Psychiatry Med 2017 May 09;52(1):3-20. [doi: 10.1177/0091217417703291]
219. Kinner EM, Armer JS, McGregor BA, Duffecy J, Leighton S, Corden ME, et al. Internet-Based Group Intervention for Ovarian Cancer Survivors: Feasibility and Preliminary Results. JMIR Cancer 2018 Jan 15;4(1):e1. [doi: 10.2196/cancer.8430]
220. Michalak EE, Morton E, Barnes SJ, Hole R, CREST.BD, Murray G. Supporting Self-Management in Bipolar Disorder: Mixed-Methods Knowledge Translation Study. JMIR Ment Health 2019 Apr 15;6(4):e13493 [FREE Full text] [doi: 10.2196/13493] [Medline: 30985287]
221. Townsend L, Zippay A, Caler K, Forenza B. Technology and Opportunity: People with Serious Mental Illness and Social Connection. Journal of the Society for Social Work and Research 2016 Jun;7(2):371-393. [doi: 10.1086/686882]
222. Muuraiskangas S, Mattila E, Kyttälä P, Koreasalo M, Lappalainen R. User experiences of a mobile mental well-being intervention among pregnant women. : Springer Verlag; 2016 Presented at: International Symposium on Pervasive Computing Paradigms for Mental Health; November 28-29; Barcelona, Spain p. 140-149. [doi: 10.1007/978-3-319-32270-4_14]
223. Connolly SL, Miller CJ, Koenig CJ, Zamora KA, Wright PB, Stanley RL, et al. Veterans' Attitudes Toward Smartphone App Use for Mental Health Care: Qualitative Study of Rurality and Age Differences. JMIR Mhealth Uhealth 2018 Aug 22;6(8):e10748. [doi: 10.2196/10748]
224. Brandt L, Hidalgo L, Diez-Canseco F, Araya R, Mohr D, Menezes P, et al. Addressing Depression Comorbid With Diabetes or Hypertension in Resource-Poor Settings: A Qualitative Study About User Perception of a Nurse-Supported Smartphone App in Peru. JMIR Ment Health 2019 Jun 18;6(6):e11701 [FREE Full text] [doi: 10.2196/11701] [Medline: 31215511]
225. East M, Havard B, Hastings N. Mental Health Mobile Apps' Instruction: Technology Adoption Theories Applied in a Mixed Methods Study of Counseling Faculty. Journal of Technology in Human Services 2016 Oct 21;34(4):301-325. [doi: 10.1080/15228835.2016.1233842]
226. Nicholas J, Huckvale K, Larsen ME, Basu A, Batterham PJ, Shaw F, et al. Issues for eHealth in Psychiatry: Results of an Expert Survey. J Med Internet Res 2017 Feb 28;19(2):e55. [doi: 10.2196/jmir.6957]

227. Krog MD, Nielsen MG, Le JV, Bro F, Christensen KS, Mygind A. Barriers and facilitators to using a web-based tool for diagnosis and monitoring of patients with depression: a qualitative study among Danish general practitioners. BMC Health Serv Res 2018 Jun 27;18(1):503 [FREE Full text] [doi: 10.1186/s12913-018-3309-1] [Medline: 29945613]
228. Lemey C, Larsen ME, Devylder J, Courtet P, Billot R, Lenca P, et al. Clinicians' Concerns About Mobile Ecological Momentary Assessment Tools Designed for Emerging Psychiatric Problems: Prospective Acceptability Assessment of the MEmind App. J Med Internet Res 2019 Apr 25;21(4):e10111. [doi: 10.2196/10111]
229. Baumel A, Schueller SM. Adjusting an Available Online Peer Support Platform in a Program to Supplement the Treatment of Perinatal Depression and Anxiety. JMIR Mental Health 2016 Mar 21;3(1):e11. [doi: 10.2196/mental.5335]
230. Navarro P, Bambling M, Sheffield J, Edirippulige S. Exploring Young People's Perceptions of the Effectiveness of Text-Based Online Counseling: Mixed Methods Pilot Study. JMIR Ment Health 2019 Jul 03;6(7):e13152. [doi: 10.2196/13152]
231. Saurman E, Perkins D, Roberts R, Roberts A, Patfield M, Lyle D. Responding to mental health emergencies: implementation of an innovative telehealth service in rural and remote new South Wales, Australia. J Emerg Nurs 2011 Sep;37(5):453-459. [doi: 10.1016/j.jen.2010.11.005] [Medline: 21889653]
232. Baumel A. Online emotional support delivered by trained volunteers: users' satisfaction and their perception of the service compared to psychotherapy. J Ment Health 2015;24(5):313-320. [doi: 10.3109/09638237.2015.1079308] [Medline: 26485198]
233. Saurman E, Kirby S, Lyle D. No longer 'flying blind': how access has changed emergency mental health care in rural and remote emergency departments, a qualitative study. BMC Health Serv Res 2015 Apr 14;15:156 [FREE Full text] [doi: 10.1186/s12913-015-0839-7] [Medline: 25889260]
234. Nicholas J, Shilton K, Schueller SM, Gray EL, Kwasny MJ, Mohr DC. The Role of Data Type and Recipient in Individuals' Perspectives on Sharing Passively Collected Smartphone Data for Mental Health: Cross-Sectional Questionnaire Study. JMIR Mhealth Uhealth 2019 Apr 05;7(4):e12578 [FREE Full text] [doi: 10.2196/12578] [Medline: 30950799]
235. Schuster L, Drennan J, Lings I. Understanding consumers' decisions to adopt technology-enabled transformative services. The Service Industries Journal 2015 Oct 14;35(15-16):846-864. [doi: 10.1080/02642069.2015.1090979]
236. Kasckow J, Zickmund S, Rotondi A, Mrkva A, Gurklis J, Chinman M, et al. Development of telehealth dialogues for monitoring suicidal patients with schizophrenia: consumer feedback. Community Ment Health J 2014 Apr;50(3):339-342 [FREE Full text] [doi: 10.1007/s10597-012-9589-8] [Medline: 23306676]
237. Vereenooghe L, Gega L, Langdon PE. Intellectual disability and computers in therapy: Views of service users and clinical psychologists. Cyberpsychology: Journal of Psychosocial Research on Cyberspace 2017 May 31;11(1):1-15. [doi: 10.5817/CP2017-1-11]
238. Jorm AF. Mental health literacy: Empowering the community to take action for better mental health. American Psychologist 2012 Apr;67(3):231-243. [doi: 10.1037/a0025957]
239. Helsper EJ, Eynon R. Distinct skill pathways to digital engagement. European Journal of Communication 2013 Sep 16;28(6):696-713. [doi: 10.1177/0267323113499113]
240. 404 FE. European Citizens' Digital Health Literacy. European Commission. 2014. URL: https://ec.europa.eu/commfrontoffice/publicopinion/flash/fl_404_en.pdf [accessed 2020-12-01]
241. Apolinário-Hagen J, Fritsche L, Bierhals C, Salewski C. Improving attitudes toward e-mental health services in the general population via psychoeducational information material: A randomized controlled trial. Internet Interv 2018 Jun;12:141-149 [FREE Full text] [doi: 10.1016/j.invent.2017.12.002] [Medline: 30135778]
242. Possemato K, Kuhn E, Johnson EM, Hoffman JE, Brooks E. Development and refinement of a clinician intervention to facilitate primary care patient use of the PTSD Coach app. Behav. Med. Pract. Policy Res 2016 Feb 22;7(1):116-126. [doi: 10.1007/s13142-016-0393-9]
243. Jamison R, Jurcik D, Edwards R, Huang C, Ross E. A Pilot Comparison of a Smartphone App With or Without 2-Way Messaging Among Chronic Pain Patients: Who Benefits From a Pain App? Clin J Pain 2017 Aug;33(8):676-686 [FREE Full text] [doi: 10.1097/AJP.0000000000000455] [Medline: 27898460]
244. Richards P, Simpson S, Bastiampillai T, Pietrabissa G, Castelnuovo G. The impact of technology on therapeutic alliance and engagement in psychotherapy: The therapist's perspective. Clinical Psychologist 2020 Nov 10;22(2):171-181 [FREE Full text] [doi: 10.1111/cp.12102]
245. Schulze N, Reuter SC, Kuchler I, Reinke B, Hinkelmann L, Stöckigt S, et al. Differences in Attitudes Toward Online Interventions in Psychiatry and Psychotherapy Between Health Care Professionals and Nonprofessionals: A Survey. Telemed J E Health 2018 Nov 09;25(10):926-932. [doi: 10.1089/tmj.2018.0225] [Medline: 30412450]
246. Simons L, Valentine AZ, Falconer CJ, Groom M, Daley D, Craven MP, et al. Developing mHealth Remote Monitoring Technology for Attention Deficit Hyperactivity Disorder: A Qualitative Study Eliciting User Priorities and Needs. JMIR mHealth uHealth 2016 Mar 23;4(1):e31. [doi: 10.2196/mhealth.5009]
247. Drozd F, Haga SM, Lisøy C, Slinning K. Evaluation of the implementation of an internet intervention in well-baby clinics: A pilot study. Internet Interventions 2018 Sep;13:1-7. [doi: 10.1016/j.invent.2018.04.003]
248. Peters D, Deady M, Glozier N, Harvey S, Calvo RA. Worker Preferences for a Mental Health App Within Male-Dominated Industries: Participatory Study. JMIR Ment Health 2018 Apr 25;5(2):e30. [doi: 10.2196/mental.8999]

249. Janz NK, Becker MH. The Health Belief Model: A Decade Later. Health Education Quarterly 2016 Sep 04;11(1):1-47. [doi: 10.1177/109019818401100101]
250. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
251. Schueller SM, Tomasino KN, Mohr DC. Integrating Human Support Into Behavioral Intervention Technologies: The Efficiency Model of Support. Clin Psychol Sci Pract 2016 Nov 17;24(1):27-45. [doi: 10.1111/cpsp.12173]
252. De Choudhury M, De S. Mental Health Discourse on reddit: Self-Disclosure, Social Support, and Anonymity Munmun. 2014 Presented at: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media; June 1–4; Ann Arbor, Michigan p. 71-80.
253. Venkatesh V, Bala H. Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Sciences 2008 May;39(2):273-315 [FREE Full text] [doi: 10.1111/j.1540-5915.2008.00192.x]
254. Pung A, Fletcher SL, Gunn JM. Mobile App Use by Primary Care Patients to Manage Their Depressive Symptoms: Qualitative Study. J Med Internet Res 2018 Sep 27;20(9):e10035 [FREE Full text] [doi: 10.2196/10035] [Medline: 30262449]
255. Santini ZI, Koyanagi A, Tyrovolas S, Mason C, Haro JM. The association between social relationships and depression: a systematic review. J Affect Disord 2015 Apr 1;175:53-65. [doi: 10.1016/j.jad.2014.12.049] [Medline: 25594512]
256. Ng MM, Firth J, Minen M, Torous J. User Engagement in Mental Health Apps: A Review of Measurement, Reporting, and Validity. Psychiatric Services 2019 Jul;70(7):538-544. [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]

Abbreviations

CBT: cognitive behavioral therapy
DMHI: digital mental health intervention

Edited by G Eysenbach; submitted 18.09.20; peer-reviewed by A Wasil, M Deady, N Andalibi, CY Lin; comments to author 07.11.20; revised version received 24.12.20; accepted 08.02.21; published 24.03.21.

Please cite as:
Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, Stadnick N, Zheng K, Mukamel D, Sorkin DH
Barriers to and Facilitators of User Engagement With Digital Mental Health Interventions: Systematic Review
J Med Internet Res 2021;23(3):e24387
URL: https://www.jmir.org/2021/3/e24387
doi:10.2196/24387
PMID:33759801

Review

The Value of Applying Ethical Principles in Telehealth Practices: Systematic Review

Amanda Jane Keenan[1], BA, Grad Cert, Grad Dip, MBA, MPH; George Tsourtos[1], BPsych, PhD; Jennifer Tieman[1], BSc, PhD

College of Medicine and Public Health, Flinders University, Bedford Park, Australia

Corresponding Author:
Amanda Jane Keenan, BA, Grad Cert, Grad Dip, MBA, MPH
College of Medicine and Public Health
Flinders University
Sturt Road
Bedford Park, 5042
Australia
Phone: 61 8 7221 8418
Email: keen0076@flinders.edu.au

Abstract

Background: As the use of technology to deliver health services is increasing rapidly and has further intensified during the COVID-19 pandemic, these initiatives may fail if ethical impacts are not fully identified and acted upon by practitioners. Ignoring the ethical impacts of information and communication technology health service delivery creates an unintended risk for patients and can lead to reduced effectiveness, noncompliance, and harm, undermining the best intentions of governments and clinicians.

Objective: Our aim was to explore how ethical considerations or impacts may be different, greater, or more variable in information and communication technology methods versus face-to-face health care delivery models, and how they may be applied in practice.

Methods: We undertook a systemic literature review to provide a critical overview of existing research into the incorporation of ethical principles into telehealth practice. Six databases were searched between March 2016 to May 2016 and again in December 2020 to provide the benefit of currency. A combination of broad terms ("ethics," "ethical," "health," and "care") with the restrictive terms of "telehealth" and "telemedicine" was used in keyword searches. Thematic analysis and synthesis of each paper was conducted, aligned to the framework developed by Beauchamp and Childress.

Results: From the 49 papers reviewed, authors identified or discussed the following ethical principles in relation to telehealth practice: autonomy (69% of authors, 34/49), professional–patient relationship (53% of authors, 26/49), nonmaleficence (41% of authors, 20/49), beneficence (39%, of authors, 19/49), and justice (39% of authors, 19/49).

Conclusions: Although a small number of studies identified ethical issues associated with telehealth practice and discussed their potential impact on service quality and effectiveness, there is limited research on how ethical principles are incorporated into clinical practice. Several studies proposed frameworks, codes of conduct, or guidelines, but there was little discussion or evidence of how these recommendations are being used to improve ethical telehealth practice.

(J Med Internet Res 2021;23(3):e25698) doi:10.2196/25698

KEYWORDS

telehealth; ethics; telemedicine; ethical; telecare; review; patient experience; care; effectiveness; framework

Introduction

This literature review provides a critical overview of the existing research into the incorporation of ethical principles into telehealth practice because in the last decade, the use of information and communication technology (ICT) to deliver health services has rapidly grown into a rich tapestry of applications. Governments have sought to reduce health care expenditure and improve efficiency and access to care by the use of telephone, video, remote monitoring, or online methods [1]. Health service responses to the recent COVID-19 pandemic have further accelerated the use of telehealth globally, increasing the need for research and effective knowledge transfer mechanisms. Evaluations of telehealth programs have focused on the promises of telehealth, such as improved efficacy, efficiency, and clinical outcomes. Less research has focused on the potential perils of telehealth and the potentially negative,

harmful, or unethical impacts of this type of service delivery [2]. The ethical considerations in ICT methods may be different, greater, or more variable as compared to face-to-face care models. The "major polarities of the medical practice," including "respect for the patient, health-care quality and humaneness, as well as aiming at matching the needs of the whole population equitably" [3] are complicated by concerns about patient autonomy, the altered nature of the professional–patient relationship, the lack of the human touch in care, and the medicalization of the home environment [3,4]. Ignoring ethical impacts of ICT health service delivery creates unintended risk for patients and can lead to reduced effectiveness, noncompliance, and harm, undermining the best intentions of governments and clinicians [5,6].

For the purposes of this review the definition of telehealth is derived from the criteria outlined by the World Health Organization (WHO) [7]. This definition of telehealth includes the following: that the purpose of telehealth practice is to provide clinical support to patients by health care professionals, that it connects users who are not in the same physical location, that it involves the use of various types of ICT, and that its purpose is to improve health outcomes. The framework for the definitions, concepts and principles of health ethics is that provided by Beauchamp and Childress [8] and include respect for autonomy, nonmaleficence, beneficence, justice, and the professional–patient relationship.

Methods

Information Sources and Search Strategy

The sources of data included the Cochrane Database of Systematic Reviews, MEDLINE, Scopus, Web of Science, PubMed, and CINAHL databases were searched from March 2016 to May 2016, and again in December 2020 to provide the benefit of currency. The Cochrane database was included to obtain relevant control trials or clinical studies. MEDLINE, CINAHL, and PubMed were chosen as comprehensive databases of peer reviewed studies from the disciplines of medicine, nursing (particularity community nursing), and allied health professions (eg, psychology) that are most commonly associated with telehealth practice. Scopus and Web of Science were also included to supplement the results with studies from the social sciences and humanities, particularly philosophy and sociology, which had the potential to provide studies from an ethical, rather than a clinical or technological perspective. Studies using both qualitative and quantitative methods were included in the search criteria.

The terms used in the keyword search were "ethics," "ethical," "health," "telehealth," "telemedicine," and "care". During the search process the use of the broad terms, "ethics," "ethical," "health," and "care," was combined with the restrictive terms of "telehealth" and "telemedicine".

Grey Literature

Extended searching of internet sites, conference abstracts, and presentations was undertaken to identify any relevant grey literature that was not uncovered in the database search. Given the pilot or proof-of-concept structure of some telehealth services, relevant material might have been available in an unpublished form.

Inclusion and Exclusion Criteria

The criteria provided by the WHO definition of telehealth was used to guide the inclusion and exclusion process at the first refinement stage.

Stage 1: Inclusion and Exclusion Criteria Used at the Abstract and Title screen

Eligible for inclusion and critical appraisal were studies that examined or discussed a relationship between health or medical ethics and the delivery of health services by connecting patients and providers in different physical locations using ICT with the purpose of improving health outcomes.

Ineligible for inclusion were studies relating to the ethical use of digital health service delivery methods that did not include an interaction with a health professional through telephonic methods, such as the use of sensors and assistive technologies in the home, or the use of mobile health apps. Studies that included both telehealth services and nontelehealth services together were included. Duplicate studies were also removed.

Stage 2: Inclusion/Exclusion Criteria Used at the Full-Text Screen

The final studies reviewed at the full-text screening stage were further required to satisfy the inclusion criteria of identifying or discussing ethical principles in relation to telehealth practice, ethical dilemmas or challenges in telehealth practice, or the ethical framework of telehealth practice.

Results

Figure 1 illustrates the flow of the citations reviewed. The PRISMA (Preferred Reporting Items for Systematic Reviews) diagram indicates that most of the citation results came from CINAHL and the Web of Science databases. The content of these databases predominantly consists of research from the nursing, allied health, and social science disciplines, as well as grey literature. PubMed and MEDLINE produced fewer results, but these citations proved to be, in some cases, more relevant to the search terms, particularly the application of biomedical ethics in practice. The Cochrane database initial search produced only 1 citation of a reviewed control trial, and this was later excluded as it did not meet the second stage inclusion criteria, suggesting few clinical control trials have been undertaken regarding health ethics and telehealth practice.

The updated search, shown in Figure 2, produced 19 papers from The Cochrane database, suggesting the need for an increase in empirical evidence in this field. A summary of the ethical themes and distribution by article is shown in Table 1.

Figure 1. Flow of the citations reviewed as part of the systematic review 1980-2016.

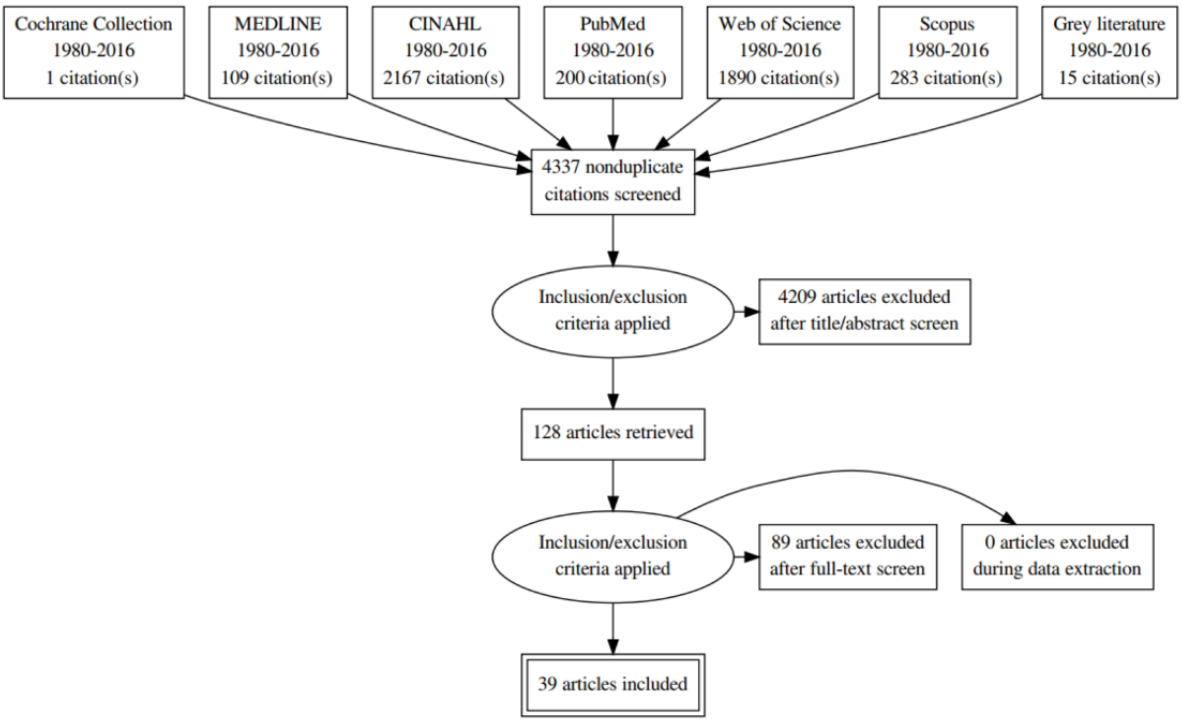


Figure 2. Flow of the citations reviewed as part of the systematic review 2017-2019.

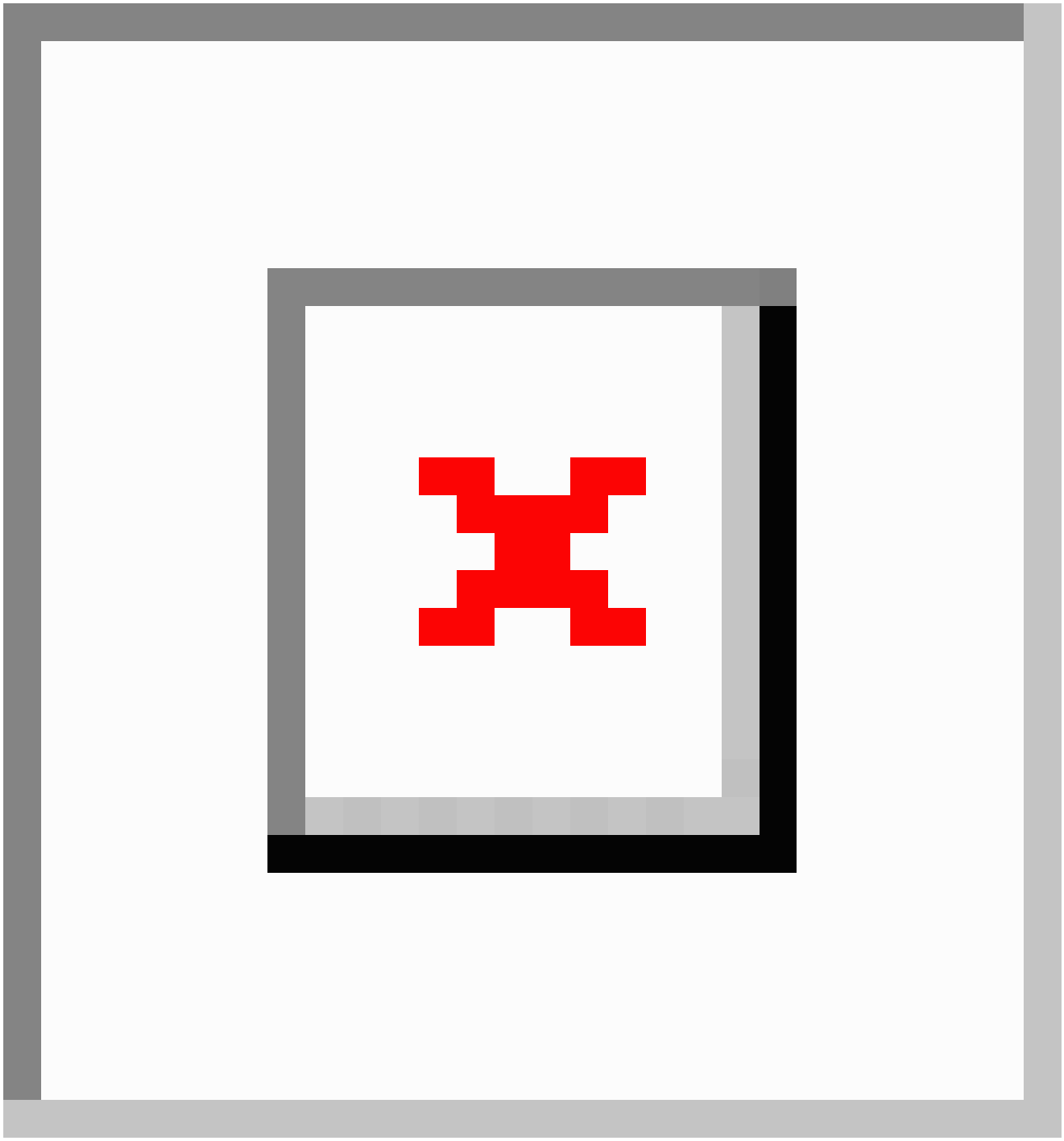


Table 1. Summary of themes and distribution of the discussion by article.

Ethical theme	Article by author and date
Autonomy	Botrugno 2019; Chaet et al 2017; Clark et al 2010; Cornford et al 2001; Demiris et al 2006; Draper et al 2013; Eccles 2010; Fisk et al 2014; Fleming 2009; Glueckauf et al 2018; Heintz et al 2015; Holmstrom et al 2007; Kaplan 2008; Korhonen 2015; Langarizadeh et al 2017; Layman 2003; Loute et al 2017; Magnusson 2003; Mort et al 2015; Nelson 2010; Nelson et al 2013; Nesher et al 2011; Newton 2014; Palm et al 2013; Parks 2016; Percival et al 2006; Perry et al 2010; Roman et al 1997; Rutenberg 2008; Sävenstedt et al 2006; Schermer 2009; Sethi et al 2012; Skar et al 2018; Sorell et al 2012; Stowe et al 2011
Beneficence	Chaet et al 2017; Clark et al 2010; Cornford et al 2002; Eccles 2010; Holmstrom et al 2008; Iserson 2000; Loute et al 2017; Magnusson 2003; Nelson et al 2013; Nesher et al 2011; Perry et al 2010; Roman et al 1997; Rutenberg 2008; Shea 2008; Skar et al 2018; Voerman et al 2017; Willems 2005;
Justice	Botrugno 2019; Chaet et al 2017; Clark et al 2010; Cornford et al 2001; Demiris et al 2009; Eccles 2010; Fleming 2009; Heintz et al 2015; Holmstrom et al2007; Humbyrd 2019; Langarizadeh et al 2017; Layman 2003; Loute et al 2017; Magnusson 2003; Nelson 2010; Nelson et al 2013; Palm et al 2013; Perry et al 2010; Skar et al 2018
Nonmaleficence	Chaet et al 2017; Clark et al 2010; Cornford et al 2001; Eccles 2010; Fleming 2009; Glueckauf et al 2018; Gogia et al 2016; Humbyrd 2019; Iserson 2000; Langarizadeh et al 2017; Loute et al 2017; Magnusson 2003; Nesher et al 2011; Perry et al 2010; Roman et al 1997; Rutenberg 2008; Sarhan 2009; Sävenstedt et al 2006; Skar et al 2018; Voerman et al 2017; Willems 2005
Professional–patient relationships	Botrugno 2019; Barina 2015; Chaet et al 2017; Cheshire 2017; Clark et al 2010; Demiris et al 2009; Draper et al 2013; Fleming 2009; Gogia et al 2016; Humbyrd 2019; Iserson 2000; Kluge 2011; Korhonen 2015; Langarizadeh et al 2017; Nelson 2010; Pols 2010; Roman et al 1997; Sävenstedt et al 2006; Skar et al 2018; Stanberry 2001; Stowe et al 2010; Voerman et al 2017; Wade et al 2012; Willems 2005

In all, 49 articles were included in the analysis stage and incorporated into a data extraction table (Multimedia Appendix 1). As the literature search did not identify any clinical studies and the number of original qualitative research studies was low at 8, a thematic analysis approach, searching across the data to "find repeated patterns of meaning" was applied [9]. The analysis and synthesis of each paper was conducted using both inductive and deductive reasoning. The papers were organized in accordance with the type of study involved and the ethical principles, frameworks, or evaluation processes identified or discussed in each one as relevant to the 5 principles of biomedical ethics. Also recorded in the data extraction table were any ethical subthemes that were present in addition to the core 5 under examination.

Among the 8 included studies that used qualitative methods to collect data, ethnographic, interview, and focus group methodologies were used. Furthermore, 6 of these studies involved patients or caregivers and nurses or other health professionals. The remaining papers were systematic reviews or research that incorporated existing literature. In addition, 6 studies recommended an ethical framework, code of conduct, or system of evaluation for the ethical provision of telehealth services. One-fifth of all papers included were from 2017 onward, indicating an increasing interest in telehealth ethics, even prior to the COVID-19 pandemic.

Discussion

Overview of Acquired Studies and Themes

The broad search strategy yielded 49 initial results, but analysis identified few studies that described how ethical considerations are or may be incorporated into telehealth practice, whether in the home, community, or medical environment. Although a small number of qualitative studies identified relevant ethical issues associated with telehealth practice and subsequently discussed their potential impact on service quality from the perspective of patients, caregivers, and health professionals, there is scant research on how ethical principles are incorporated into telehealth practice [10-18]. Several studies proposed ethical frameworks, codes of conduct, or guidelines for telehealth service delivery that may be applied or followed by health professionals, but they provided little discussion, evidence, or evaluation of how these recommendations are being used to establish or improve ethical telehealth practice [6,15,19-22].

Autonomy

Autonomy was the predominant ethical principle discussed in the literature, with 69% (34/49) of the authors identifying or discussing it in relation to telehealth practice. Within this primary theme, several subthemes emerged including consent, individual choice, independence, empowerment, control, and self-determination [23-25]. Two qualitative studies in Sweden found that autonomy can be both improved and diminished through the use of telehealth by increasing the freedom for older persons to remain living in their own homes, while also potentially contributing to their isolation and "being made captive" in their homes [16]. This issue of telehealth seeking to improve autonomy but actually having the opposite effect was noted in a UK study which found that, while the introduction of telehealth as part of a home telecare service for older patients can "drastically improve their autonomy," it may also lead to an increased reluctance to move out of the home environment for even a small amount of time and thus reduce independence [10].

A qualitative study of telenurses identified issues relating to gender-specific and cultural concerns affecting autonomy and independence specific to females accessing care [17]. A further Swedish study involving patients and families found accessing education, information, and support at a time convenient to patients could increase autonomy and a sense of independence [14].

Recommendations for maintaining or improving autonomy in telehealth practice recognize that the concepts of choice and independence are not simple, particularly for older or more vulnerable patients, and decisions about what improves autonomy "takes place in a complex and changing context" [26]. Heintz et al reduce the concept of autonomy to the patient's ability to give informed consent or participate fully in decision making; meanwhile, Palm recommends an ethical assessment design comprising 5 questions relating to patient autonomy, including co-design, behavioral adjustments, understanding of the system and control under different usage scenarios, whether it enhances independence, and if so, whether this independence is desirable [19,27]. Although a reduction in autonomy may be unavoidable for some telehealth patients, particularly in older users who are more accepting of "traditional" health care models, wherever possible, the "loss should be minimised" [15,28]. Layman [29] notes that the methods of data collection, storage, and manipulation used with telehealth may threaten patient autonomy if it becomes the primary source of information, and recommends a "multipronged approach" in incorporating ethical principles into practice, including regulations, standards, codes of conduct, and codes of ethics. The implications for ethical telehealth practice from the perspective of autonomy then are that care should be taken to robustly assess the impact on patients from a number of standpoints to reduce the potential risk [30].

Beneficence

Analysis revealed that 39% (19/49) of the papers identified or discussed the ethical principle of beneficence, or "being disposed to act for the benefit of others" [8] in relation to telehealth practice. These authors all noted that telehealth has the potential to benefit people by providing assurance, increasing an individual's confidence in managing their health, and reducing the dependence on professional caregivers or family [20,31,32]. Improving access, quality of health care availability, and the continuity of care are additional examples of telehealth increasing beneficence [31], as is the ability of patients to be treated in familiar surroundings rather than hospitals [33]. Although Beauchamp and Childress [8] note that "obligations to confer benefits can be linked to the goal of morality itself" and are an "implicit assumption" in the actions of medical professionals, the principle of beneficence informs rather than determines or justifies other moral principles. Thus, an ethical telehealth practitioner is one who provides information that empowers patients to act in their own best interest, and the wide availability of the telephone system in the majority of countries offers a greater capacity for the patients to control their own care [22,34]. From the perspective of families and caregivers, Magnusson [14] found that the use of telehealth can deliver beneficence by providing them with "education, information and support which would directly help them in their individual caring situation". In developing and implementing telehealth policies and guidelines then, it may enhance practice to be able to clearly articulate the benefits to both patients and providers in design and delivery, so that telehealth remains "a support system for well-defined needs and not be pushed as an engineering solution to health" [35].

Nonmaleficence

Our analysis further revealed that 41% (20/49) of the papers identified or discussed the ethical principle of nonmaleficence, or preventing harm, in relation to telehealth practice. Examples of telehealth's ability to actively promote safety were identified, including telephone or video lines left open for providers to check on a patient at regular intervals acting as a security guarantee against harm occurring in the home, or the mode of delivery lowering the risk in patient care because of the lack of physical proximity of the health care worker to the patient [14,22]. The potential for harm is more prevalent, however, and includes telehealth equipment such as videophones situated in the home having the effect of stigmatizing a person and causing shame or embarrassment, the possibility that professional caregivers may choose remote communication rather than delivering care in person in difficult or high needs cases may put clients at risk, and an "undue burden" [10] being imposed on unwell or frail patients who find the technology intrusive or do not fully understand its use [12]. A ethnographic study in the Netherlands with nurses and their patients found that "the feeling of safety and security the patients experienced, may not always have been realistic" due to nurses having to make value judgements about the types of information that were most important during telehealth sessions [11].

Sarhan [36] links confidentiality, nonmaleficence, and the professional responsibility of practitioners to ensure patients are protected from "emotional, spiritual, social or material" harm, while Willems [33] notes that using telehealth instead of traditional methods of health care may lead to families and caregivers being "loaded with more and different responsibilities". Nesher [37] suggests that the additional layers of technology may compromise patient care by adding complexity and obscuring the most important information from clinicians. The responsibility to "respect, preserve and defend the patient's dignity" has also been identified and linked to person-centered practice and user-driven design as core to ethical telehealth services [38]. A recent study of psychologist's telebehavioral health practices noted that over half of the survey respondents reported "inadequate skills in managing crisis situations in the context of online practice," including managing suicide risk [18]. The implications for practice here are that potential harms are not straightforward or easy to discern and may not be captured in established procedures or service evaluation tools.

Justice

Results indicated that 39% (19/49) of papers identified or discussed the ethical principle of justice in relation to telehealth practice. Justice was most discussed in relation to fairness concerning equal access to telehealth technology balancing the needs of the individual with those of the wider community, ensuring not to disadvantage one group in favor of another [10]. Examples are given where the key advantage for providing telehealth—access to care for marginalized communities—is negated by the affordability of the technology or creates additional barriers for "at risk" patients [39,40]. In the case of mental health services, Nelson et al [41] note that the criteria set by mental health professionals of only using high-standard

equipment can impact the ability of some localities to make telehealth services available, while Nesher [37] suggests that locations most likely to be in need of and benefit from telehealth services—rural areas—are likely to be least able to afford them. Perry et al [10] note the distinction between "individual level" and "system level" equity, arguing that benefits derived from the use of telehealth can positively impact in other areas of social care. Demiris [42] points out that providing underserved older adults access to services should not be done solely as a cost-saving exercise that "deprives patients of face-to-face consultations," while Fleming [43] argues that special skills in telehealth delivery should be developed to ensure access for older patients in nursing homes—the "underserved"—and ethnic minorities. When considering justice in relation to developing a telehealth practice, questions related to equal access and fair distribution of the technology, and whether a digital or information divide exists should be used to guide the implementation of telehealth services in practice [27]. Models should be evaluated, not just in terms of resource allocation but also in relation to "the principle of human value" as well as any current legislation against discrimination [19]. Botrugno [44], in discussing the argument for telehealth to underpin greater distributive justice in health care, advises against accepting "technological determinism," arguing instead for a "plan of analysis through which to critically assess the implications of telehealth" [44].

Professional–Patient Relationships

Our review further found that 53% (26/49) of the papers identified or discussed the potential disruption of the relationship between health professionals and their patients, with several subthemes emerging including confidentiality, privacy, and fidelity [21,45-47]. The lack of the "human touch" in care has been identified as a key concern in providing health services remotely although the importance of this may vary between disciplines, such as teledermatology where it may be low and telepsychology where it may be much greater [42]. As more health services are delivered in "the virtual realm" rather than in physical proximity, the risk increases of "creating a distance between touch and care" [48]. Fleming [43] suggests that telehealth should not be used to replace the traditional face-to-face methods of health care delivery "that remains crucial to healing" but rather should be viewed as a supplementary method to improve care and treatment.

The undermining of trust between patients and their health care providers was discussed within the ethical subtheme of fidelity, with Chaet [49] asserting that the practice of medicine is "inherently a moral activity, founded in a covenant of trust between patient and physician," which must be sustained. Trust and mutual respect may be challenged between patients and providers in a telehealth environment, particularly if the two have never met in person [50], as through "words and nonverbal actions the patient and the physician establish a relationship of trust that is essential to good medical care" [51]. The notion of not just trust but "sound trust" has been raised in relation to telehealth, whereby additional actions or behaviors are required by health professionals to win public trust "in the face of the conflicting interests that are at stake" [52]. Trust may also be undermined by the skepticism or caution generated by unfamiliar equipment being directed towards the health professional or by the reluctance of patients to speak freely in the presence of such equipment due to privacy or communication concerns [53,54]. Savenstedt [12] links the use of technology to the notion of "superficial care" arising from the "superficial" relationship created by the replacement of face-to-face care with remote care, but also notes that communicating through the use of telehealth may reduce loneliness in people who otherwise would have few options for interaction. Finally, Wade et al [13] found that patients may in some cases find a telehealth communication setting more protected and feel they are more likely to be listened to by the health provider, but also suggested that palliative care patients may suffer through a lower-quality therapeutic relationship, as may "groups or families". The implication is that care should be taken around context and patient preferences for the relationship with telehealth practitioners when designing services.

Limitations

Limitations noted during the implementation of the search strategy were the following: broad search terms such as "ethics" and "ethical" might have resulted in missing relevant papers that used similar but different terms, such as "moral," "virtue/virtues," or "values"; manual searches of journals specifically dedicated to studies on the use of ICT in health care might have provided additional suitable studies for inclusion; and several studies that were identified for inclusion at the first stage were not able to be obtained in their entirety and consequently could not be assessed for the second stage or included in the results.

Conclusions

Our findings suggest that the principles of biomedical ethics are relevant to the context of telehealth practice and that interest in how ethical principles impact telehealth service delivery for patients and clinicians is increasing. We have identified a number of considerations for future telehealth policy and practice development to reduce the risk to patient experience and improve clinical care and delivery effectiveness and sustainability. Further research into how ethical principles are incorporated into organizational telehealth policies and models of care documentation would identify how ethical priorities are aligned with care delivery in current practice. Investigation and analysis of how ethical principles are incorporated into telehealth practice from both a patient and provider experience would identify gaps and opportunities to develop purposeful frameworks and guidelines, supported by an appropriate knowledge transfer model for telehealth clinicians.

Acknowledgments

AK was funded by an Australian Postgraduate Award.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Table 2 Data Extraction.
[DOC File , 100 KB - jmir_v23i3e25698_app1.doc]

References

1. Telemedicine: Opportunites and developments in member states. Report on the second global survey on ehealth. World Health Organization. 2010. URL: https://www.who.int/goe/publications/goe_telemedicine_2010.pdf [accessed 2021-09-07]
2. Thede LQ. Telehealth: promise or peril? Online J Issues Nurs 2001;6(3):1. [Medline: 11936940]
3. Loute A, Cobbaut J. What ethics for telemedicine? In: The Digitization of Healthcare: New Challenges and Opportunities. London: Palgrave Macmillan; 2017:399-416.
4. Schermer M. Telecare and self-management: opportunity to change the paradigm? J Med Ethics 2009 Nov;35(11):688-691. [doi: 10.1136/jme.2009.030973] [Medline: 19880706]
5. Kaplan B, Litewka S. Ethical challenges of telemedicine and telehealth. Camb Q Healthc Ethics 2008 Oct 01;17(4):401-416. [doi: 10.1017/s0963180108080535]
6. Cornford T, Klecun-Dabrowska E. Ethical perspectives in evaluation of telehealth. Camb Q Healthc Ethics 2001;10(2):161-169. [doi: 10.1017/s0963180101002079] [Medline: 11302093]
7. A health telematics policy in support of WHO's Health-For-All strategy for global health development: report of the WHO group consultation on health telematics, 11-16 December, Geneva. World Health Organization. 1998. URL: http://apps.who.int/iris/bitstream/10665/63857/1/WHO_DGO_98.1.pdf [accessed 2021-07-02]
8. Beauchamp T, Childress J. Priciples of Biomedical Ethics. 7th ed. United States of America: Oxford University Press; 2013.
9. Liamputtong P. Qualitative Research Methods. 4th ed. Australia: Oxford University Press; 2013.
10. Perry J, Beyer S, Francis J, Holmes P. At a glance 24: ethical issues in the use of telecare. UK. London: Social Care Institute for Excellence URL: https://www.scie.org.uk/publications/ataglance/ataglance24.asp [accessed 2018-06-08]
11. Pols J. The heart of the matter. About good nursing and telecare. Health Care Anal 2010 Jan 10;18(4):374-388. [doi: 10.1007/s10728-009-0140-1] [Medline: 20063198]
12. Sävenstedt S, Sandman P, Zingmark K. The duality in using information and communication technology in elder care. J Adv Nurs 2006 Oct;56(1):17-25. [doi: 10.1111/j.1365-2648.2006.03975.x] [Medline: 16972914]
13. Wade VA, Eliott JA, Hiller JE. A qualitative study of ethical, medico-legal and clinical governance matters in Australian telehealth services. J Telemed Telecare 2012 Jan 20;18(2):109-114. [doi: 10.1258/jtt.2011.110808]
14. Magnusson L, Hanson E. Ethical issues arising from a research, technology and development project to support frail older people and their family carers at home. Health & Social Care in the Community 2003;11(5):431-439. [doi: 10.1046/j.1365-2524.2003.00446.x] [Medline: 14498840]
15. Mort M, Roberts C, Pols J, Domenech M, Moser I. Ethical implications of home telecare for older people: a framework derived from a multisited participative study. Health Expect 2015 Jun;18(3):438-449. [doi: 10.1111/hex.12109] [Medline: 23914810]
16. Percival J, Hanson J. Big brother or brave new world? Telecare and its implications for older people's independence and social inclusion. Critical Social Policy 2016 Jun 29;26(4):888-909. [doi: 10.1177/0261018306068480]
17. Holmström I, Höglund AT. The faceless encounter: ethical dilemmas in telephone nursing. J Clin Nurs 2007 Oct;16(10):1865-1871. [doi: 10.1111/j.1365-2702.2007.01839.x]
18. Glueckauf RL, Maheu MM, Drude KP, Wells BA, Wang Y, Gustafson DJ, et al. Survey of psychologists' telebehavioral health practices: technology use, ethical issues, and training needs. Professional Psychology: Research and Practice 2018 Jun;49(3):205-219. [doi: 10.1037/pro0000188]
19. Heintz E, Lintamo L, Hultcrantz M, Jacobson S, Levi R, Munthe C, et al. Framework for systematic identification of ethical aspects of healthcare technologies: the SBU approach. Int J Technol Assess Health Care 2015 Jan;31(3):124-130 [FREE Full text] [doi: 10.1017/S0266462315000264] [Medline: 26134927]
20. Iserson KV. Telemedicine: a proposal for an ethical code. Camb Q Healthc Ethics 2000;9(3):404-406. [doi: 10.1017/s0963180100003133] [Medline: 10858890]
21. Langarizadeh M, Moghbeli F, Aliabadi A. Application of ethics for providing telemedicine services and information technology. Med Arch 2017 Oct;71(5):351-355 [FREE Full text] [doi: 10.5455/medarh.2017.71.351-355] [Medline: 29284905]
22. Roman L, Stockwell S, Louvier V, Clements F. Creating an ethical foundation for home telemedicine. Home Health Care Management & Practice 2016 Jun 30;9(6):58-66. [doi: 10.1177/108482239700900612]

23. Demiris G, Oliver D, Courtney K. Ethical considerations for the utilization of tele-health technologies in home and hospice care by the nursing profession. Nurs Adm Q 2006;30(1):56-66. [doi: 10.1097/00006216-200601000-00009] [Medline: 16449885]
24. Parks JA. Home-based care, technology, and the maintenance of selves. HEC Forum 2015 Jun;27(2):127-141. [doi: 10.1007/s10730-015-9278-4] [Medline: 25787720]
25. Sethi R, Bagga G, Carpenter D, Azzi D, Khusainov R. Telecare: legal, ethical and socioeconomic factors. London: Legal, ethical and socioeconomic factors. International Conference on Telehealth (Telehealth 2012); 2012 Presented at: IET Seminar on Assisted Living 2011; 6th April 2011; London URL: https://researchportal.port.ac.uk/portal/files/156094/Sethi_et_al_Telecare-Legal_Ethical_and_Socioeconomic_Factors_IASTED_2012.pdf [doi: 10.1049/ic.2011.0023]
26. Draper H, Sorell T. Telecare, remote monitoring and care. Bioethics 2013 Sep;27(7):365-372. [doi: 10.1111/j.1467-8519.2012.01961.x] [Medline: 22500571]
27. Palm E, Nordgren A, Verweij M, Collste G. Ethically sound technology? Guidelines for interactive ethical assessment of personal health monitoring. Stud Health Technol Inform 2013;187:105-114. [Medline: 23920461]
28. Fisk M, Rudel D. Telehealth and Service Delivery in the Home: Care, Support, and the Importance of User Autonomy. In: George C, Whitehouse D, Duquenoy P, editors. eHealth: Legal, Ethical and Governance Challenges. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013:211-225.
29. Layman E. Health informatics: ethical issues. Health Care Manag (Frederick) 2003;22(1):2-15. [doi: 10.1097/00126450-200301000-00002] [Medline: 12688606]
30. Sorell T, Draper H. Telecare, surveillance, and the welfare state. Am J Bioeth 2012;12(9):36-44. [doi: 10.1080/15265161.2012.699137] [Medline: 22881854]
31. Clark P, Capuzzi K, Harrison J. Telemedicine: medical, legal and ethical perspectives. Med Sci Monit 2010 Dec;16(12):RA261-RA272. [Medline: 21119593]
32. Eccles A. Ethical considerations around the implementation of telecare technologies. Journal of Technology in Human Services 2010 May 10;28(1-2):44-59. [doi: 10.1080/15228831003770759]
33. Willems D. Advanced home care technology: moral questions associated with an ethical ideal. Health Council of the Netherlands Centre for Ethics and Health. 2005. URL: https://www.healthcouncil.nl/documents/advisory-reports/2004/06/29/advanced-home-care-technology-moral-questions-associated-with-an-ethical-ideal [accessed 2018-06-08]
34. Rutenberg C, Oberle K. Ethics in telehealth nursing practice. Home Health Care Management & Practice 2008 Feb 21;20(4):342-348. [doi: 10.1177/1084822307310766] [Medline: 105775613]
35. Gogia SB, Maeder A, Mars M, Hartvigsen G, Basu A, Abbott P. Unintended consequences of tele health and their possible solutions. contribution of the IMIA working group on telehealth. Yearb Med Inform 2016 Nov 10(1):41-46 [FREE Full text] [doi: 10.15265/IY-2016-012] [Medline: 27830229]
36. Sarhan F. Telemedicine in healthcare. 2: The legal and ethical aspects of using new technology. Nurs Times 2009;105(43):18-20. [Medline: 19950459]
37. Nesher L, Jotkowitz A. Ethical issues in the development of tele-ICUs. J Med Ethics 2011 Nov;37(11):655-657. [doi: 10.1136/jme.2010.040311] [Medline: 21546522]
38. Skär L, Söderberg S. The importance of ethical aspects when implementing eHealth services in healthcare: A discussion paper. J Adv Nurs 2018 May;74(5):1043-1050. [doi: 10.1111/jan.13493] [Medline: 29105151]
39. Humbyrd CJ. Virtue ethics in a value-driven world: ethical telemedicine. Clin Orthop Relat Res 2019 Dec;477(12):2639-2641 [FREE Full text] [doi: 10.1097/CORR.0000000000000908] [Medline: 31764325]
40. Newton M. The promise of telemedicine. Surv Ophthalmol 2014;59(5):559-567. [doi: 10.1016/j.survophthal.2014.02.003] [Medline: 25130894]
41. Nelson E, Davis K, Velasquez S. 4 - Ethical Considerations in Providing Mental Health Services Over Videoteleconferencing. In: Myers K, Turvey C, editors. Telemental Health. USA: Elsevier Inc; 2013:47-62.
42. Demiris G, Doorenbos AZ, Towle C. Ethical considerations regarding the use of technology for older adults. The case of telehealth. Res Gerontol Nurs 2009 Apr;2(2):128-136. [doi: 10.3928/19404921-20090401-02] [Medline: 20077974]
43. Fleming D, Edison K, Pak H. Telehealth ethics. Telemed J E Health 2009 Oct;15(8):797-803. [doi: 10.1089/tmj.2009.0035] [Medline: 19780693]
44. Botrugno C. Towards an ethics for telehealth. Nurs Ethics 2017 May 15;26(2):357-367. [doi: 10.1177/0969733017705004]
45. Kluge EW. Ethical and legal challenges for health telematics in a global world: Telehealth and the technological imperative. International Journal of Medical Informatics 2011 Feb;80(2):e1-e5. [doi: 10.1016/j.ijmedinf.2010.10.002]
46. Korhonen E, Nordman T, Eriksson K. Technology and its ethics in nursing and caring journals. Nurs Ethics 2014 Oct 20;22(5):561-576. [doi: 10.1177/0969733014549881]
47. Stanberry B. Telemedicine: barriers and opportunities in the 21st century. J Intern Med 2000 Jun;247(6):615-628 [FREE Full text] [doi: 10.1046/j.1365-2796.2000.00699.x] [Medline: 10886483]
48. Barina R. New places and ethical spaces: philosophical considerations for health care ethics outside of the hospital. HEC Forum 2015 Jun;27(2):93-106. [doi: 10.1007/s10730-015-9277-5] [Medline: 25796213]

49. Chaet D, Clearfield R, Sabin J, Skimming K, Council on Ethical Judicial Affairs American Medical Association. Ethical practice in telehealth and telemedicine. J Gen Intern Med 2017 Oct;32(10):1136-1140 [FREE Full text] [doi: 10.1007/s11606-017-4082-2] [Medline: 28653233]
50. Nelson WA. The ethics of telemedicine. Unique nature of virtual encounters calls for special sensitivities. Healthc Exec 2010;25(6):50-53. [Medline: 21229905]
51. Cheshire W. Telemedicine and the ethics of medical care at a distance. Ethics and Medicine 2017;33(2):71-75 [FREE Full text]
52. Voerman SA, Nickel PJ. Sound trust and the ethics of telecare. J Med Philos 2017 Feb;42(1):33-49. [doi: 10.1093/jmp/jhw035] [Medline: 27932397]
53. SHEA K, EFFKEN JA. Enhancing patients' trust in the virtual home healthcare nurse. CIN: Computers, Informatics, Nursing 2008;26(3):135-141. [doi: 10.1097/01.ncn.0000304782.34694.12]
54. Stowe S, Harding S. Telecare, telehealth and telemedicine. European Geriatric Medicine 2010 Jun;1(3):193-197. [doi: 10.1016/j.eurger.2010.04.002]

Abbreviations

ICT: information and communication technology
PRISMA: Preferred Reporting Items for Systematic Reviews
WHO: World Health Organization

Edited by G Eysenbach; submitted 11.11.20; peer-reviewed by Z Aghaei, A Khaleghi, H Mehdizadeh, J op den Buijs, MA Bahrami; comments to author 02.12.20; revised version received 04.01.21; accepted 15.03.21; published 30.03.21.

Please cite as:
Keenan AJ, Tsourtos G, Tieman J
The Value of Applying Ethical Principles in Telehealth Practices: Systematic Review
J Med Internet Res 2021;23(3):e25698
URL: https://www.jmir.org/2021/3/e25698
doi:10.2196/25698
PMID:33783366

Review

Blended Self-Management Interventions to Reduce Disease Burden in Patients With Chronic Obstructive Pulmonary Disease and Asthma: Systematic Review and Meta-analysis

Xiaoyue Song[1], MSc; Cynthia Hallensleben[1], BSc; Weihong Zhang[2], PhD; Zongliang Jiang[2], BSc; Hongxia Shen[1], MSc; Robbert J J Gobbens[3,4,5], PhD; Rianne M J J Van Der Kleij[1], PhD; Niels H Chavannes[1], MD, PhD; Anke Versluis[1], PhD

[1]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands

[2]Faculty of Nursing and Health, Zhengzhou University, Zhengzhou, China

[3]Faculty of Health, Sports and Social Work, Inholland University of Applied Sciences, Amsterdam, Netherlands

[4]Zonnehuisgroep Amstelland, Amstelveen, Netherlands

[5]Department Family Medicine and Population Health, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium

Corresponding Author:
Xiaoyue Song, MSc
Department of Public Health and Primary Care
Leiden University Medical Center
Albinusdreef 2
Leiden, 2333 ZA
Netherlands
Phone: 31 613836216
Email: x.song@lumc.nl

Abstract

Background: Chronic obstructive pulmonary disease (COPD) and asthma have a high prevalence and disease burden. Blended self-management interventions, which combine eHealth with face-to-face interventions, can help reduce the disease burden.

Objective: This systematic review and meta-analysis aims to examine the effectiveness of blended self-management interventions on health-related effectiveness and process outcomes for people with COPD or asthma.

Methods: PubMed, Web of Science, COCHRANE Library, Emcare, and Embase were searched in December 2018 and updated in November 2020. Study quality was assessed using the Cochrane risk of bias (ROB) 2 tool and the Grading of Recommendations, Assessment, Development, and Evaluation.

Results: A total of 15 COPD and 7 asthma randomized controlled trials were included in this study. The meta-analysis of COPD studies found that the blended intervention showed a small improvement in exercise capacity (standardized mean difference [SMD] 0.48; 95% CI 0.10-0.85) and a significant improvement in the quality of life (QoL; SMD 0.81; 95% CI 0.11-1.51). Blended intervention also reduced the admission rate (relative ratio [RR] 0.61; 95% CI 0.38-0.97). In the COPD systematic review, regarding the exacerbation frequency, both studies found that the intervention reduced exacerbation frequency (RR 0.38; 95% CI 0.26-0.56). A large effect was found on BMI (d=0.81; 95% CI 0.25-1.34); however, the effect was inconclusive because only 1 study was included. Regarding medication adherence, 2 of 3 studies found a moderate effect (d=0.73; 95% CI 0.50-0.96), and 1 study reported a mixed effect. Regarding self-management ability, 1 study reported a large effect (d=1.15; 95% CI 0.66-1.62), and no effect was reported in that study. No effect was found on other process outcomes. The meta-analysis of asthma studies found that blended intervention had a small improvement in lung function (SMD 0.40; 95% CI 0.18-0.62) and QoL (SMD 0.36; 95% CI 0.21-0.50) and a moderate improvement in asthma control (SMD 0.67; 95% CI 0.40-0.93). A large effect was found on BMI (d=1.42; 95% CI 0.28-2.42) and exercise capacity (d=1.50; 95% CI 0.35-2.50); however, 1 study was included per outcome. There was no effect on other outcomes. Furthermore, the majority of the 22 studies showed some concerns about the ROB, and the quality of evidence varied.

Conclusions: In patients with COPD, the blended self-management interventions had mixed effects on health-related outcomes, with the strongest evidence found for exercise capacity, QoL, and admission rate. Furthermore, the review suggested that the interventions resulted in small effects on lung function and QoL and a moderate effect on asthma control in patients with asthma.

There is some evidence for the effectiveness of blended self-management interventions for patients with COPD and asthma; however, more research is needed.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42019119894; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=119894

(J Med Internet Res 2021;23(3):e24602) doi:10.2196/24602

KEYWORDS

blended intervention; COPD; asthma; meta-analysis; systematic review

Introduction

Background

Chronic lung diseases are the leading cause of disability and death worldwide [1]. Of all chronic lung diseases, chronic obstructive pulmonary disease (COPD) and asthma are the most prevalent [1]. There were approximately 251 million cases of COPD globally in 2015, and COPD is predicted to become the third leading cause of death by 2030 [2]. Approximately 300 million people have asthma worldwide, with a projected increase of an additional 100 million people by 2025 [3]. The impact of a health problem, measured by financial cost, morbidity, and other indicators, is called disease burden. It is often quantified in terms of disability-adjusted life years (DALYs) or quality-adjusted life years (QALYs) [1]. In 2017, the loss of DALYs was the first for COPD and the second for asthma [1]. In addition, a loss in health-related quality of life (QoL) is seen in many patients (eg, a decline in health, increased hospital admissions, and high medication costs). The World Health Organization estimates that the cost of a QALY for COPD ranges from US $6700 to $13,400 due to exacerbations and medication. In patients with asthma, annual costs vary from less than US $150 to US $3000 [4,5]. There is increased awareness that self-management represents a promising strategy to decrease disease burden [6]. Self-management could improve patient outcomes and decrease disease burden by supporting patients to positively adapt their health behaviors and develop skills to better manage their diseases [7].

Self-management refers to an individual's ability to manage their symptoms, treatment, physical and psychosocial consequences, and lifestyle changes inherent to life with a chronic condition [8]. In traditional face-to-face self-management interventions, patients with COPD and asthma are equipped with the knowledge and skills to manage their health condition successfully [9]. Previous studies have found that these self-management interventions are effective in improving disease knowledge and self-efficacy [10]. However, these face-to-face self-management interventions are limited by their accessibility (eg, lower accessibility for patients who are more distant to the health care provider or when the health care provider lacks time) [11].

eHealth is an alternative to traditional face-to-face interventions. The most cited definition of eHealth is "health services and information delivered or enhanced through the internet and related technologies" [12]. Compared with traditional face-to-face interventions, eHealth interventions can be cost and time saving and offer better accessibility and flexibility [13]. Moreover, eHealth interventions can help optimize the therapeutic process, increase treatment efficiency, and decrease costs by enhancing (web-based) communication possibilities between health care providers and patients [14]. There have been promising results with eHealth self-management interventions [15,16]. A meta-analysis showed that, for patients with COPD, eHealth self-management programs (eg, web-based phone calls and web-based interventions) led to a significant improvement in symptoms [15]. However, eHealth interventions typically allow for limited tailoring of patients' needs and lower patient engagement [17]. There have also been concerns about reliability, security, confidentiality, and lack of education and training [18]. These factors can negatively impact the implementation and effectiveness of these interventions.

The most recent development is the blended intervention. There are different definitions of blended interventions [19,20]. We use the definition by Erbe et al [20]: "Treatment programs that use elements of both face-to-face and internet-based interventions, including both the integrated and the sequential use of both treatment formats." Blended interventions could retain the positive aspects of face-to-face interventions and eHealth by mitigating their negative aspects. Furthermore, blended intervention could diminish the number of face-to-face contacts needed and provide support that is available at all times [21]. With eHealth, patients can also monitor their health condition throughout the day and convey their health information to health care providers without time and distance limitations. Patients can also receive quick assistance during critical periods of care facilitated by real-time alerts and reminders, which could help patients adhere to their action plan. For patients with COPD and asthma, blended interventions can include various elements [22,23] (eg, training, education, and action plans) with different blended intervention components (eg, internet-based phone calls and individual face-to-face interventions, web platforms combined with individual face-to-face interventions) [22,23]. Some studies have shown that blended self-management interventions are effective in improving QoL in patients with COPD and asthma [24,25].

Current reviews suggest that blended interventions could be effective [19,20], but these reviews are limited for several reasons. First, the reviews focus on mental health and not on chronic lung diseases [20]. Second, the reviews focus on health-related effectiveness outcomes and not on process outcomes [19]. Third, the reviews do not specifically focus on self-management interventions [19,20]. To conclude, a comprehensive overview or meta-analysis of the effect of blended self-management interventions on the disease burden

of patients with COPD and asthma, including process outcomes and health-related effectiveness outcomes, is lacking.

Objectives

A systematic review will be performed to assess the effectiveness of blended self-management interventions in patients with COPD and asthma. When appropriate, a meta-analysis will be conducted. Internet-based, telephone, and SMS-delivered interventions are included because all of these are parts of eHealth [13]. Thus, this study aims to investigate the effectiveness of blended self-management interventions in patients with COPD and asthma.

Methods

Systematic Review and Meta-analysis

This review follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [26]. The review was registered in PROSPERO (number 2019: CRD42019119894).

Search Strategy

A search strategy was established in collaboration with a certified librarian to identify relevant studies on blended self-management interventions in patients with COPD and asthma. A total of 5 electronic databases (ie, PubMed, Web of Science, COCHRANE Library, Emcare, and Embase) were searched on December 28, 2018, and updated on November 30, 2020. There were search terms related to 4 areas: (1) COPD or asthma, (2) eHealth, (3) face-to-face intervention, and (4) blended intervention (Multimedia Appendix 1). The search terms related to COPD or asthma and blended interventions were first combined, resulting in 84 studies. Due to the limited number of studies, the search terms associated with COPD or asthma were combined with terms about eHealth and face-to-face interventions. In every database, the search was limited to peer-reviewed publications. The search strategy was not restricted based on publication year, as we aimed to provide a comprehensive overview of the use of blended interventions in patients with COPD and asthma. In addition, reference lists of the included studies and previous reviews were searched to identify additional studies that might be eligible for inclusion.

Eligibility Criteria

The patient, intervention, comparison, outcome, study design tool was used to develop an effective search strategy and determine the inclusion and exclusion criteria [27]. The following inclusion criteria were used to identify the studies: (1) participants: adults (≥18 years old) with COPD or asthma; (2) intervention: blended self-management intervention (consisting of an eHealth component combined with a face-to-face component); (3) comparison: eHealth intervention with or without usual care (UC) and face-to-face intervention with or without UC or only UC; (4) outcome measures: health-related effectiveness or process outcomes; and (5) individual randomized controlled trials (RCTs). Studies were excluded if: (1) the participants were children or adolescents, (2) the eHealth apps were only used to collect data, (3) outcomes were not about the health-related outcomes, and (4) studies were cluster RCTs.

Study Selection

After the removal of duplicates, the identified titles and abstracts were screened for eligibility. If insufficient information was provided, the full-text paper was screened. When a full-text paper was not available, a request was sent to the authors. Studies that did not meet the inclusion criteria were excluded. Screening the titles, abstracts, and full texts was performed by 2 reviewers independently (XS and ZJ). Any disagreements between the 2 authors were resolved by a third reviewer (CH).

Data Collection and Coding

Data were collected using a standardized data extraction form. It included (1) study characteristics (eg, first author, publication year, country, number and age of patients, percentage of female patients, disease severity or diagnosis, setting [ie, home, primary care (PC), or secondary care (SC)]), intervention, and follow-up duration), (2) intervention characteristics (ie, category and functionality of the eHealth and face-to-face component), (3) behavior change techniques (BCTs) used in the blended self-management intervention, and (4) the health-related effectiveness and process outcomes. Information was extracted from each publication by XS and ZJ. Inter-rater reliability, as assessed with Cohen κ, indicated strong agreement (κ=0.90) [28].

COPD severity was classified based on the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria [29]. Patients were considered to have COPD when the ratio between forced expiratory volume in 1s (FEV1) and full forced vital capacity (FVC) was <0.70. The degree of obstruction was defined as follows: (1) GOLD I: FEV1 ≥80% predicted (mild), (2) GOLD II: 50%≤FEV1<80% predicted (moderate), (3) GOLD III: 30%≤FEV1<50% predicted (severe), and (4) GOLD IV: FEV1 <30% predicted (very severe). There is no standard classification of severity for patients with asthma.

As mentioned above, different intervention characteristics were extracted from the publications. First, the eHealth component of the intervention was categorized as a mobile app; eg, phone call or SMS), an internet-assisted intervention (eg, web page, chat room), or multiple component interventions with multiple eHealth technologies. Second, the function of the eHealth app was categorized into informing, instructing, displaying, guiding, reminding or alerts, and communicating (ie, between provider and patients) [30]. Third, face-to-face interventions were classified as individual (eg, home visits, PC or SC visits) or group-based interventions (eg, group pulmonary rehabilitation). Fourth, the function of the face-to-face intervention was classified as (1) education: introduction of disease-related information and how to use eHealth, (2) training: provide information about self-management, (3) consultation: discuss individual action plan, (4) assessment: test and assess the patient's performance, or (5) monitoring: provide reminders to improve intervention adherence [31,32].

Outcome indicators were classified into health-related effectiveness outcome or process outcome indicators. Health-related effectiveness outcome indicators included

outcomes related to disease status and health condition (ie, exercise capacity, dyspnea, lung function, QoL, admission, mortality, exacerbation frequency, and BMI). Process outcome indicators included intermediate outcomes during the implementation process (eg, visits, satisfaction, costs, smoking, self-management ability, physical activity, medication and therapy adherence, psychosocial, symptom management, nutrition, and alcohol). A positive effect was ascribed when there was a significant positive effect of the intervention on the outcome measure compared with the control group (CG). When the outcome measure did not significantly differ between the intervention group (IG) and CG, it was rated as *no effect*. A mixed effect was ascribed when multiple measures were used to measure a similar outcome, and the effect on the measures was in different directions (eg, in the study by Garcia [22], there was a significant positive effect on inhaler treatment adherence, whereas there was no effect on oral treatment adherence).

Quality Assessment

Study quality was assessed using the Cochrane risk of bias (ROB) 2 tool [33]. The tool assessed 5 domains of potential bias: (1) randomization, (2) deviations from the intended interventions (effect of assignment to intervention), (3) missing outcome data, (4) measurement of the outcome, and (5) selection of the reported result. Each domain had a few signaling questions. On the basis of the authors' (XS and ZJ) responses to the signaling questions, a judgment on the ROB (*low, some concerns*, or *high*) for each domain could be made to assess the bias that might confound the study findings [33]. The quality of the clinical evidence was critically appraised using the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) system [34], which evaluated the risk for bias, inconsistency, indirectness, and imprecision for each outcome. Four categories were used to define the quality of evidence: high quality of evidence (the true effect lies close to that of the effect estimate), moderate quality of evidence (the true effect is likely to be close to the effect estimate, but there is a possibility that it is substantially different), low quality of evidence (the true effect may be substantially different from the effect estimate), and very low quality of evidence (the true effect is likely to be substantially different from the effect estimate) [35]. The quality assessment was performed by XS and ZJ, and any disagreements were resolved through discussion. Inter-rater reliability, as assessed with Cohen κ [28], indicated that there was strong agreement between raters (κ=0.80).

Data Analysis

When an outcome was assessed using different measurements in one study, data from the most specific disease-related questionnaire were used. For example, in the study by Garcia [22], QoL was measured using both the Saint-George's Respiratory Questionnaire (SGRQ), a specific QoL questionnaire, and Euroqol, a generic health-related QoL questionnaire. SGRQ was selected and analyzed in the meta-analysis because it is the most specific disease-related questionnaire.

First, a systematic review was conducted to determine the results. For continuous data, Cohen *d* was recommended to calculate the effect size [36] (Cohen *d* >0.2=small effect, Cohen *d* >0.5=moderate effect, and Cohen *d* >0.8=large effect) [37]. For dichotomous data, the relative ratio (RR) was calculated to assess the effect size. An RR greater than 1 indicates an increased likelihood that the stated outcome is achieved in the IG. If the RR is less than 1, there is a decreased likelihood that the outcome is achieved in the IG. A ratio of 1 indicated no difference (ie, the outcome was just as likely to occur in the IG as it was in the CG) [38].

When 3 or more studies reported on the same outcome measure, this outcome was included in the meta-analysis [39]. For continuous data, the standardized mean difference (SMD) accounted for the same outcomes measured with different assessment tools (eg, QoL was assessed using the SGRQ, COPD assessment test [CAT], and chronic respiratory questionnaire [CRQ]). SMDs were used to standardize the results of the studies to a uniform scale before they could be combined in the quantitative synthesis. SMDs and associated 95% CIs were used to calculate the mean difference and SD difference between the IG and CG for each study. When the mean or SD was not mentioned, the author was contacted for missing information. Cohen *d* was used to interpret the data [37]. For dichotomous data, RR was calculated to assess the effect size [38]. Publication bias was tested if more than 10 studies report on the same outcome measure [40]. $P<.05$ was considered significant for the effect estimate.

A random-effect model was used because the variance of study populations and intervention designs was anticipated as heterogeneity across the included studies [41]. Heterogeneity was assessed using chi-square tests and I^2 statistics [42]. A P value of <.1 indicates statistically significant heterogeneity. The I^2 statistic was used to quantify the size of the heterogeneity between studies: 25%, 50%, and 75% can be considered small, medium, and substantial heterogeneity, respectively [42]. Outliers were identified using the value of the standardized residual [43]. Studies whose standardized residual was equal to or larger than 1.96 were identified as an outlier and were excluded from the meta-analysis. No subgroup analysis was planned because of the limited number of studies. All analyses were performed using the Review Manager (RevMan version 5.4; The Cochrane Collaboration) and Stata version 14.0 (StataCorp) [44].

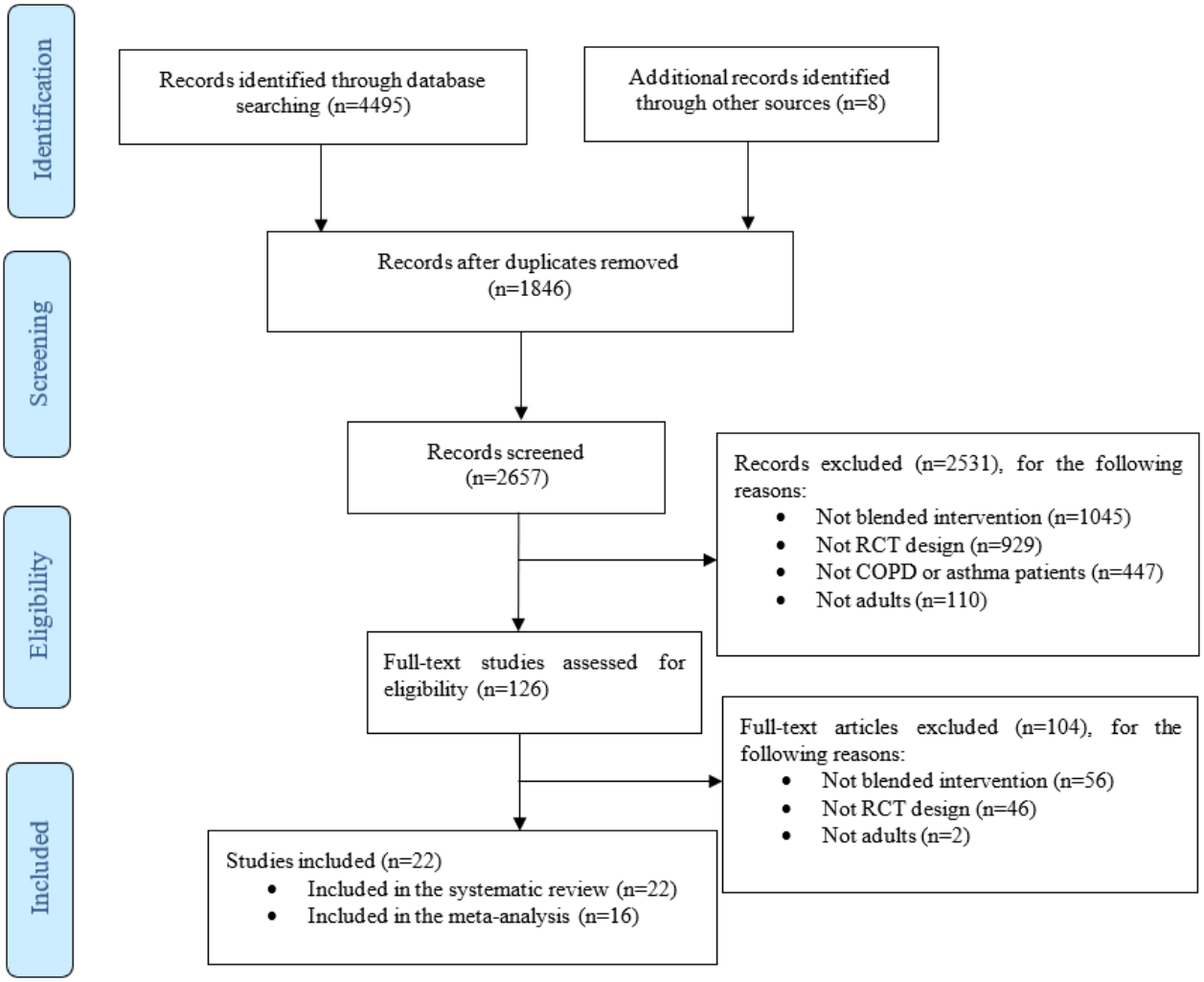
Results

Search Results

The literature search identified a total of 4495 potentially eligible records, and 2657 records remained after duplicates were excluded. After screening the titles and abstracts, additional 2531 records were excluded for other reasons (Figure 1). The full texts of the remaining 126 studies were assessed, and 22 RCTs [22-25,45-62] were included in this review. Of the 22 RCTs, 2 were pilot RCT studies [45,51] and 1 was a feasibility RCT [48]. These studies were included because they followed the CONSORT (Consolidated Standards of Reporting Trials) checklist [45,51], and they were small sample size RCTs [48,51]. A total of 15 RCTs focused on patients with COPD [22,24,45-57]. Of these studies, 11 were included in the

meta-analysis [22,24,45,47,48,50,51,54-57]. The remaining 4 studies [46,49,52,53] were excluded because no available means and SDs were reported or obtained after contacting the authors. A total of 7 studies focused on patients with asthma [23,25,58-62]. Of the 7 asthma studies with available data, 5 were pooled into a meta-analysis [25,58,60-62]. The other 2 studies were not included in the meta-analysis because of the lack of means and SDs after contacting the authors.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) flowchart of the systematic review and meta-analysis. COPD: chronic obstructive pulmonary disease; RCT: randomized controlled trial.



Study Characteristics

A total of 15 COPD studies [22,24,45-57] were published between 2006 and 2020 and were conducted in China (n=5) [48,54-57], United States (n=2) [24,51], Denmark (n=2) [49,52], Canada (n=1) [53], England (n=1) [45], Spain (n=1) [22], Germany (n=1) [50], Australia (n=1) [46], and 1 in both Spain and Belgium [47]. The sample size ranged from 39 to 242 (with a total sample size of 1477). The average age of patients with COPD ranged from 64.10 to 73.50 years. Of the 15 COPD studies, 8 had UC as a CG [22,24,46,47,49,54,56,57], 5 had a *visit* as CG (meaning that the health care provider visited the patients' home or patients visited the PC or SC) [45,48,51,52,54], and 2 studies had both UC and visits in the CG [50,53]. The settings were home and SC (n=9) [22,24,46,47,53-57], home care (n=2) [45,48], and home care and PC (n=4) [49-52]. The duration of the blended self-management interventions ranged from 4 to 48 weeks, with a mean of 22.13 weeks (SD 16.20). The follow-up duration ranged from 17 to 48 weeks.

Seven asthma studies [23,25,58-63] were published from 2003 to 2020 and were conducted in the Netherlands (n=3) [25,61,62], Germany (n=1) [59], England (n=1) [23], United States (n=1) [60], and China (n=1) [58]. The study sample size ranged from 16 to 200 (total N=527). The mean age of patients with asthma ranged from 24.80 to 52.00 years. CG included UC (n=4) [23,25,60,62] and visits (n=3) [58,59,61]. The duration of the blended self-management interventions ranged from 3 to 48 weeks, with a mean of 15.88 weeks (SD 13.48). The follow-up duration ranged from 36 to 120 weeks. An overview of the study characteristics is provided in Table 1.

Table 1. Study characteristics of chronic obstructive pulmonary disease and asthma studies.

COPD[a] and asthma study[b]	Country	Participants		Setting	Participants, mean (SD)		Gender (female), n (%)	Severity[c]	CG[d]	Intervention (weeks)	Follow-up (weeks)
		IG[e]	CG		IG	CG					
COPD (included in the meta-analysis)											
Bentley et al [45]	England	25	23	Home	67.20 (11.60)	65.90 (9.40)	—[f]	—	Home visits	8	32
Chau et al [48]	China	22	18	Home	73.50 (6.10)	72.20 (6.10)	1 (3)	II-IV	Home visits	8	—
Casas et al [47]	Spain and Belgium	65	90	Home and SC[g]	70.00 (90.00)	72.00 (90.00)	26 (16.8)	I-IV	UC[h]	4	48
Garcia et al [22]	Spain	21	41	Home and SC	73.00 (6.00)	74.00 (8.00)	8 (13)	—	UC	48	—
Jehn et al [50]	Germany	32	30	Home and PC[i]	64.10 (10.90)	69.10 (9.20)	14 (23)	II-IV	UC+PC visits	36	—
Koff et al [24]	United States	20	20	Home and SC	66.60 (9.10)	65.00 (8.20)	21 (53)	III-IV	UC	12	—
Nguyen et al [51]	United States	19	20	Home and PC	68.00 (8.30)	70.90 (8.60)	17 (44)	—	Home visits	24	—
Wang et al [54]	China	55	65	Home and SC	69.30 (7.80)	71.90 (8.10)	63 (52.5)	II-IV	UC	24	48
Wang et al [55]	China	39	39	Home and SC	63.20 (7.50)	64.40 (7.00)	23 (30)	Mostly II-IV	SC visits	48	—
Wei et al [56]	China	42	45	Home and SC	65.20 (8.10)	63.90 (6.20)	—	I-IV	UC	24	48
Xin et al [57]	China	114	113	Home and SC	64.20 (14.20)	64.60 (14.50)	141 (62.1)	—	UC	48	—
COPD (not included in the meta-analysis)											
Cameron et al [46]	Australia	35	30	Home and SC	68.00 (9.90)	70.00 (6.80)	—	I-IV	UC	8	17
Haesum et al [49]	Denmark	47	43	Home and PC	70.20 (9.00)	69.50 (10.10)	47 (52)	I-IV	UC	4	40
Sorknaes et al [52]	Denmark	121	121	Home and PC	71.00 (10.00)	72.00 (9.00)	—	I-IV	PC visits	12	26
Stamenova et al [53]	Canada	41	41	Home and SC	71.98 (9.52)	71.76 (7.28)	36 (44)	II-IV	SC visits	24	—
Stamenova et al [53]	Canada	41	40	Home and SC	71.98 (9.52)	72.78 (9.16)	37 (46)	II-IV	UC	24	—
Asthma (included in the meta-analysis)											
Cao et al [58]	China	37	30	Home and SC	39.10 (14.30)	41.40 (12.00)	52 (78)	—	SC visits	12	—
Ostojic et al [60]	United States	8	8	Home and PC	24.80 (6.30)	24.50 (7.00)	7 (44)	M[j]	UC	16	—
Türk et al [61]	The Netherlands	7	10	SC	41.57 (12.54)	41.90 (8.58)	13 (77)	—	SC visits	12	48
Türk et al [61]	The Netherlands	14	10	SC	41.57 (9.73)	41.90 (8.58)	19 (79)	—	SC visits	12	48
van der Meer et al [25]	The Netherlands	101	99	Home and SC	36.00 (19.00; 50.00)	37.00 (18.00; 50.00)	139 (69.5)	—	UC	12	36
van Gaalen et al [62]	The Netherlands	47	60	Home and SC	36.00 (8.70)	37.00 (8.00)	76 (71.0)	—	UC	48	120

COPD[a] and asthma study[b]	Country	Participants		Setting	Participants, mean (SD)		Gender (female), n (%)	Severity[c]	CG[d]	Intervention (weeks)	Follow-up (weeks)
		IG[e]	CG		IG	CG					
Asthma (not included in the meta-analysis)											
Barbanel et al [23]	England	12	12	Home and SC	45.00 (17.00)	47.00 (17.00)	13 (54)	—	UC	12	—
Kohler et al [59]	Germany	41	41	Home and PC	49.00 (12.00)	52.00 (8.00)	32 (39)	—	PC[i] visits	3	—

[a]COPD: chronic obstructive pulmonary disease.

[b]Study by Bentley et al [45] and Nguyen et al [51] were feasibility RCTs, and study by Chau et al [48] was a pilot RCT. There was 1 study including 1 intervention group and 2 control groups (study by Stamenova et al [53]). Study by Türk et al [61] included 2 intervention groups and 1 control group.

[c]COPD severity was classified according to GOLD (Global Initiative for Chronic Obstructive Lung Disease) classification. Asthma severity was classified by the physician diagnosis.

[d]CG: control group.

[e]IG: intervention group.

[f]Not reported in the study.

[g]SC: secondary care.

[h]UC: usual care.

[i]PC: primary care.

[j]M: moderate severity.

Quality Assessment

Methodological Quality

The ROB is summarized in Table 2. Among the 15 COPD studies, the overall ROB was rated as *some concerns* in 10 studies [22,46,47,49,52-57] and *high* in 5 studies [24,45,48,50,51]. In addition, 2 studies had some concerns in the randomization process [48,50], and 13 studies showed a low ROB in the randomization process [22,24,45-47,49,51-57]. The majority of the studies showed some concerns [22,45-47,49,51-57], whereas 3 studies showed high risk from intended intervention [24,48,50]. A low ROB due to missing outcome data was found in 14 studies [22,24,45,46,48-57], whereas 1 study showed some concerns [47]. The ROB in the measurement of the outcome had some concerns in 13 studies [22,24,45,46,48-51,53-57] and a low ROB in 2 studies [47,52]. A low ROB in the selection of the reported result was found in the majority of studies [22,24,46-50,52-57], and 2 studies had some concerns [45,51].

In asthma studies, the overall ROB indicated some concerns in 4 studies [23,25,61,62] and high risk in 3 studies [58-60]. Four studies showed a low ROB in the randomization process [23,25,61,62], and 3 studies showed some concerns [58-60]. All studies indicated some concerns due to deviations from the intended intervention [23,25,58-62]. In total, 6 studies showed a low ROB outcome data [23,25,59-62], and 1 study had some concerns due to missing outcome data [58]. All studies showed some concerns in the measurement of the outcomes and low ROB in the selection of the reported results [23,25,58-62].

Table 2. Risk of bias judgments for chronic obstructive pulmonary disease and asthma randomized controlled trials.

COPD[a] or asthma study	Bias arising from the randomization process	Bias due to deviations from the intended intervention	Bias due to missing outcome data	Bias in measurement of the outcome	Bias in selection of the reported result	Overall bias
COPD						
Bentley et al [45]	L[b]	S[c]	L	S	S	H[d]
Cameron et al [46]	L	S	L	S	L	S
Casas et al [47]	L	S	S	L	L	S
Chau et al [48]	S	H	L	S	L	H
Garcia [22]	L	S	L	S	L	S
Haesum et al [49]	L	S	L	S	L	S
Jehn et al [50]	S	H	L	S	L	H
Koff et al [24]	L	H	L	S	L	H
Nguyen et al [51]	L	S	L	S	S	H
Sorknaes et al [52]	L	S	L	L	L	S
Stamenova et al [53]	L	S	L	S	L	S
Wang et al [54]	L	S	L	S	L	S
Wang et al [55]	L	S	L	S	L	S
Wei et al [56]	L	S	L	S	L	S
Xin et al [57]	L	S	L	S	L	S
Asthma						
Barbanel et al [23]	L	S	L	S	L	S
Cao et al [58]	S	S	S	S	L	H
Kohler et al [59]	S	S	L	S	L	H
Ostojic et al [60]	S	S	L	S	L	H
Türk et al [61]	L	S	L	S	L	S
van der Meer et al [25]	L	S	L	S	L	S
van Gaalen et al [62]	L	S	L	S	L	S

[a]COPD: chronic obstructive pulmonary disease.

[b]L: low risk of bias.

[c]S: some concerns.

[d]H: high risk of bias.

Quality of Evidence

In COPD studies, 19 different outcome measures were included (ie, exercise capacity, dyspnea, lung function, QoL, admission rate, exacerbation frequency, mortality, BMI, visits, satisfaction, costs, smoking, medication adherence, self-management ability, physical activity, psychosocial, symptom management, nutrition, and alcohol). Two outcome measures were rated as high quality of evidence (ie, exercise capacity and mortality), 1 measure had a moderate quality of evidence (ie, admission rate), 6 had a low quality of evidence (ie, dyspnea, lung function, QoL, visits, satisfaction, and physical activity), and the other 10 showed very low quality of evidence (exacerbation frequency, BMI, adherence, self-management ability, smoking, costs, psychosocial, symptom management, nutrition, and alcohol). In asthma studies, 10 different outcome measures were included (ie, admission rate, BMI, exercise capacity, asthma control, lung function, QoL, asthma knowledge, adherence, visits, and exacerbation frequency). Of the 10 outcomes, 7 were rated as having very low quality of evidence (ie, admission rate, BMI, exercise capacity, asthma knowledge, adherence, visits, and exacerbation frequency). Asthma control, lung function, and QoL were rated as having moderate quality of evidence (Multimedia Appendix 2 [22-25,45-62]).

Intervention Characteristic

Category of the Blended Self-Management Intervention

In COPD studies, 5 blended self-management intervention combinations were discussed: (1) multiple component eHealth and an individual face-to-face intervention (n=6) [49-53,57], (2) internet-assisted intervention and an individual face-to-face intervention (n=5) [22,45,47,48,54], (3) multiple component plus an individual and group face-to-face intervention (n=1)

[49], (4) mobile applications and an individual face-to-face intervention (n=2) [55,56], and (5) mobile applications and an individual plus group face-to-face intervention (n=1) [46].

In asthma studies, 2 blended self-management intervention combinations were discussed: (1) mobile application and individual face-to-face intervention (n=3) [23,58,60] and (2) internet-assisted intervention and the group face-to-face intervention (n=4) [25,59,61,62]. Detailed information on the interventions in the COPD and asthma studies is shown in Table 3.

Table 3. Description of the blended self-management interventions in chronic obstructive pulmonary disease and asthma studies.

COPD[a] and asthma study	eHealth		Face-to-face	
	Category (details)	Functionality	Category (details)	Functionality
COPD (included in the meta-analysis)				
Bentley et al [45]	IA[b] (telehealth-supported service)	Guide, remind, and record	Individual (home visits)	Training
Chau et al [48]	IA (peripheral devices+mobile phone)	Guide, record, remind, and display	Individual (home visits)	Education and consultation
Garcia et al [22]	IA (web-based call centre)	Guide, remind, and record	Individual (SC[c] and home visits)	Assessment, education, and consultation
Jehn et al [50]	MC[d] (peripheral devices+mobile)	Display, record, and remind	Individual (outpatient visits)	Training, and monitoring
Koff et al [24]	MC (peripheral devices+web platform+phone call)	Record, display, instruct, guide, remind, and communication	Individual (home visits)	Education, consultation; training, and assessment
Nguyen et al [51]	MC (web modules+PDA[e])	Guide, remind, record, and communication	Individual (home and PC[f] visits)	Education, training, and assessment
Stamenova et al [53]	MC (peripheral devices+web platform+phone call)	Display, record, remind, guide, and communication	Individual (SC visits)	Assessment and consultation
Wang et al [54]	IA (web platform)	Guide, record, instruct, and communication	Individual (SC visit)	Monitoring
Wang et al [55]	MA[g] (web-based app)	Guide and communication	Individual (SC visits)	Education
Wei et al [56]	MA (phone call)	Guide, remind, record, and communication	Individual (PC visits)	Education, training, and assessment
Xin et al [57]	MC (phone call+web platform)	Guide, record, instruct, and communication	Individual (SC visits)	Education and training
COPD (not included in the meta-analysis)				
Cameron et al [46]	MA (phone call)	Guide and communication	Individual+group (exercise guidance)	Education and consultation
Casas et al [47]	IA (web-based app)	Display and record	Individual (SC and home visits)	Assessment, education, and consultation
Haesum et al [49]	MC (peripheral devices+web platform)	Guide, record, remind, and communication	Individual+group visits	Training and monitoring
Sorknaes et al [52]	MC (peripheral devices+web platform)	Guide, instruct, and communication	Individual (PC visits)	Consultation
Asthma (included in the meta-analysis)				
Cao et al [58]	MA (Wechat app)	Guide, remind, and communication	Individual (SC visit)	Education
Ostojic et al [60]	MA (SMS)	Guide, display, record, and communication	Individual (PC visits)	Education
Türk et al [61]	IA (web platform)	Instruct, record, and communication	Group (unclear)	Education and training
van der Meer et al [25]	IA (web platform)	Guide, remind, record, and communication	Group (unclear)	Assessment and education
van Gaalen et al [62]	IA (web platform)	Guide, remind, and communication	Group (unclear)	Education and consultation
Asthma (not included in the meta-analysis)				
Barbanel et al [23]	MA (phone call)	Guide, remind, and record	Individual (unclear)	Education

COPD[a] and asthma study	eHealth		Face-to-face	
	Category (details)	Functionality	Category (details)	Functionality
Kohler et al [59]	IA (web platform)	Guide, record, and communication	Group (unclear)	Education and training

[a]COPD: chronic obstructive pulmonary disease.

[b]IA: internet-assisted.

[c]SC: secondary care.

[d]MC: multiple component.

[e]PDA: personal digital assistant.

[f]PC: primary care.

[g]MA: mobile application.

BCTs of the Blended Self-Management Intervention

In COPD studies, the number of BCTs used in the interventions ranged from 3 to 10, with a mean of 6.42 (SD 1.99). *General information*, *Provide feedback on performance*, *Prompt self-monitoring/tracking*, and *Problem-solving/barrier* were included in 15 studies [22,24,45-57]. *Action planning* [22,46,47,51-54,56,57] and *Motivational approach* [22,24,46,47,50-52,54,55] were included in 9 studies, respectively. *Prompt review of behavioural goals* were included in 7 studies [22,46,47,49,51,53,54]. *Goal setting* was used in 6 studies [22,46,47,51,53,54]. *Social support* was reported in 4 studies [22,47,51,55], and *Emotional control training* was used in 2 studies [46,51].

In asthma studies, the number of BCTs ranged from 4 to 10, with a mean of 6.29 (SD 2.63). *General information*, *Prompt self-monitoring/ tracking*, and *Problem-solving/barrier* were used in all 7 studies [23,25,58-62]. *Provide feedback on performance* was used in 6 studies [25,58-62]. *Action planning* and *Motivational approach* were used in 4 studies [23,25,61,62]. *Goal setting* and *Prompt review of behavioural goals* were used in 3 studies [25,61,62], *Social support* was used in 2 studies [61,62], and *Emotional control training* was used in 1 study [61] (Multimedia Appendix 3 [22-25,45-62])

Effects of the Interventions

Systematic Review

In COPD studies, the following 3 health-related effectiveness outcomes were reported: mortality [45,47,52], exacerbation frequency [50,57], and BMI [22]. Regarding outcome mortality, none of the 3 studies reported any effect [45,47,52]. Regarding outcome exacerbation frequency, both studies [50,57] found that the blended self-management intervention reduced the exacerbation frequency (RR=0.38; 95% CI 0.26-0.56). A study on BMI reported that blended self-management intervention had a significant effect on BMI (d=0.81; 95% CI 0.25-1.34) [22]. Moreover, 11 different process outcomes were studied: number of visits (including home visits, PC visits, and SC visits; n=3) [47,48,50], satisfaction with the intervention (n=3) [22,24,48], medication adherence (n=3) [22,56,57], costs (n=2) [24,45], smoking (n=2) [22,46], self-management ability (n=2) [51,55], physical activity (n=2) [22,51], nutrition (n=1) [46], alcohol (n=1) [46], psychosocial management (n=1) [46], and symptom management (n=1) [46]. Of the 3 studies, 2 showed a moderate effect (d=0.73; 95% CI 0.50-0.96) [56,57], whereas the other study reported a mixed effect on medication adherence [22]. Regarding the outcome self-management ability, 1 reported a large effect (d=1.15; 95% CI 0.66-1.62) [55], and the other study showed no effect [51]. No effect was found on the other process outcome indicators. In asthma studies, 4 health-related effectiveness outcomes were reported: admission rate [60], BMI [61], exercise capacity [61], and exacerbation frequency [25]. No effect was found on the admission rate and exacerbation frequency. A large effect was found in BMI (d=1.42; 95% CI 0.28-2.42) and exercise capacity (d=1.50; 95% CI 0.35-2.50). Three process outcomes were reported: asthma knowledge (n=2) [25,59], visits (n=2) [25,60], and adherence (therapy and medication adherence; n=2) [25,60]. No effect was found on any of the process outcome indicators.

Meta-analysis

A total of 11 studies focusing on patients with COPD were included in the meta-analysis [22,24,45,48,50,51,53-57]. The following health-related effectiveness outcomes were included: exercise capacity, dyspnea, lung function, QoL, and admission rate. Three studies reported walking distance as an indicator of exercise capacity [50,51,54]. Blended self-management intervention showed a small effect on the walking distance without significant heterogeneity (SMD=0.48; 95% CI 0.10-0.85, $\chi^2_{2=}$3.3; P=.20; I^2=39%; Figure 2). No study was identified as an outlier. Dyspnea was reported in 4 studies [22,48,51,54]. It was measured using the dyspnea subscale of the CRQ [48,51], Medical Research Council [22], and the Modified Medical Research Council [54]. Lung function was measured with FEV1% [48,50,54] and FEV1/FVC (%) [22] in 4 studies. No significant difference was found in dyspnea and lung function between the IG and CG (Figure 2). No study was identified as an outlier. QoL was reported in 8 studies with SGRQ [22,24,45,54], CAT [50,55,57], and CRQ [51]. A large effect was found on QoL, with substantial heterogeneity (SMD=0.81; 95% CI 0.11-1.51; χ^2_7=108.4; P<.001; I^2=94% Figure 3). The standardized residual identified 1 study as an outlier [22]. Removal of this study resulted in an increased effect size without decreasing heterogeneity (SMD=0.90; 95% CI 0.15-1.65; χ^2_6=94.1; P<.001; I^2=94%). Furthermore, blended self-management intervention reduced admission rate with a substantial heterogeneity (RR=0.61; 95% CI 0.38-0.97;

χ^2_5=17.6; *P*=.003; I^2=72%; Figure 4). No outliers were identified.

A total of 5 asthma studies were pooled in the meta-analysis [25,58,60-62]. In addition, 3 health-related effectiveness outcomes were included: lung function, QoL, and asthma control. Lung function was reported as FEV1 (%) [58,61] and FEV_1 [25]. Blended self-management intervention showed a small effect on the lung function without significant heterogeneity (SMD=0.40; 95% CI 0.18-0.62; χ^2_4=1.5; *P*=.83; I^2=0%). No study was identified as an outlier. Three studies reported QoL using an asthma QoL questionnaire [25,58,62]. There was a small effect size of the blended self-management intervention on QoL without significant heterogeneity (SMD=0.36; 95% CI 0.21-0.50; χ^2_2=0.8; *P*=.68; I^2=0%). No study was identified as an outlier. Furthermore, 3 studies reported asthma control using an asthma control questionnaire [25,58,62]. A moderate effect was found in the blended intervention self-management group without significant heterogeneity (SMD=0.67; 95% CI 0.40-0.93; χ^2_2=3.0; *P*=.23; I^2=33%; Figure 5). No study was identified as an outlier.

Figure 2. Forest plots for (A) exercise capacity, (B) dyspnea, and (C) lung function in chronic obstructive pulmonary disease studies.

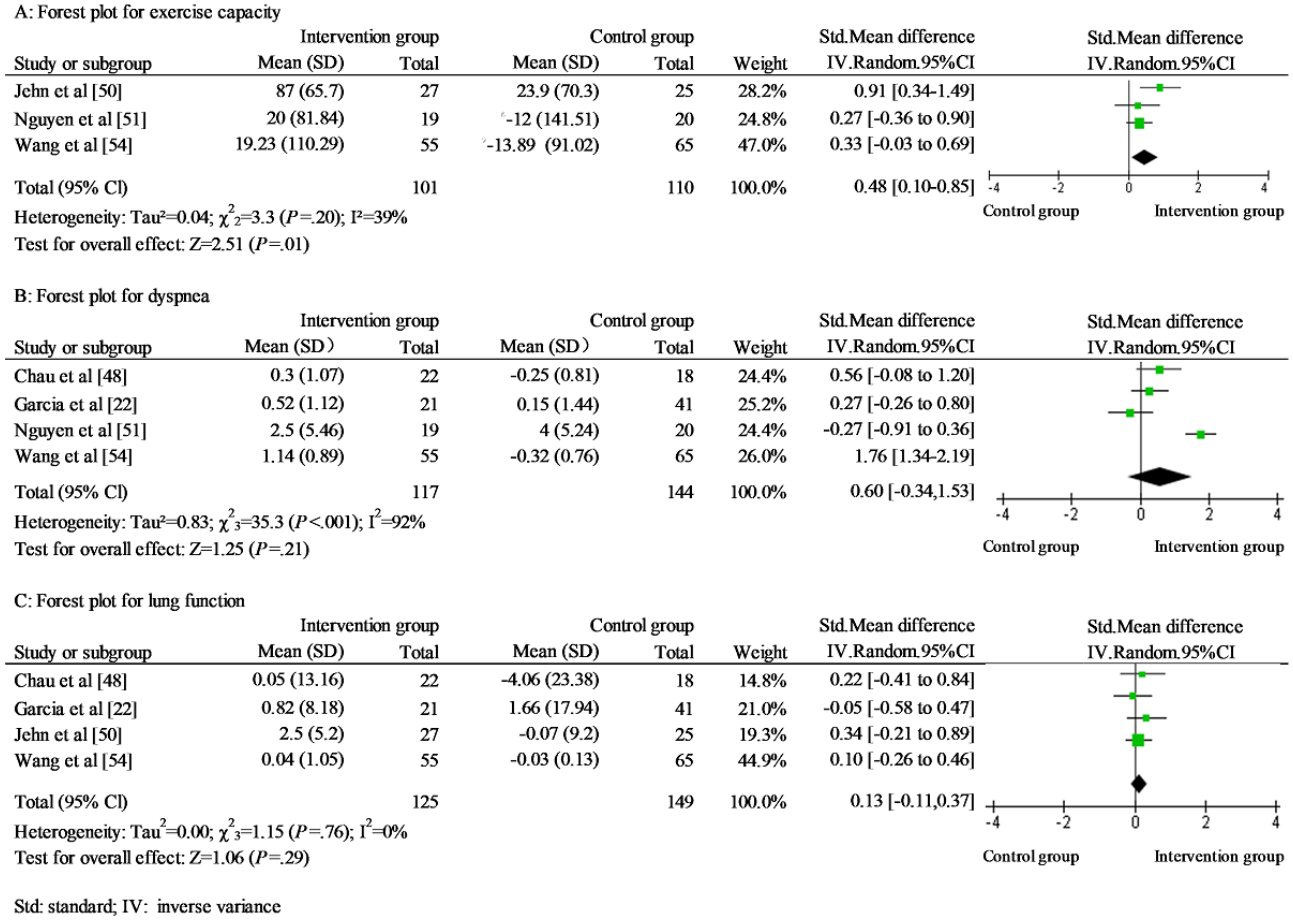


Figure 3. Forest plot for quality of life in chronic obstructive pulmonary disease studies.

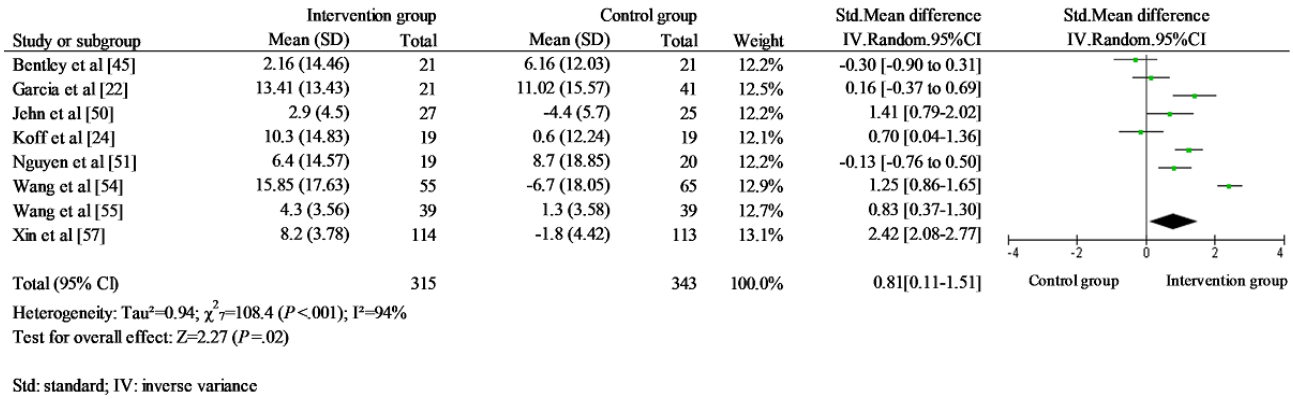


Figure 4. Forest plot for admission rate in chronic obstructive pulmonary disease studies.

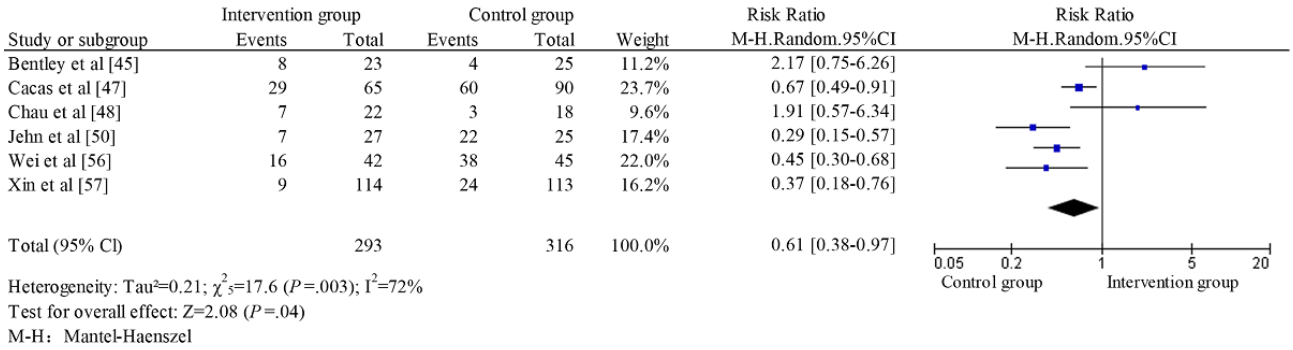
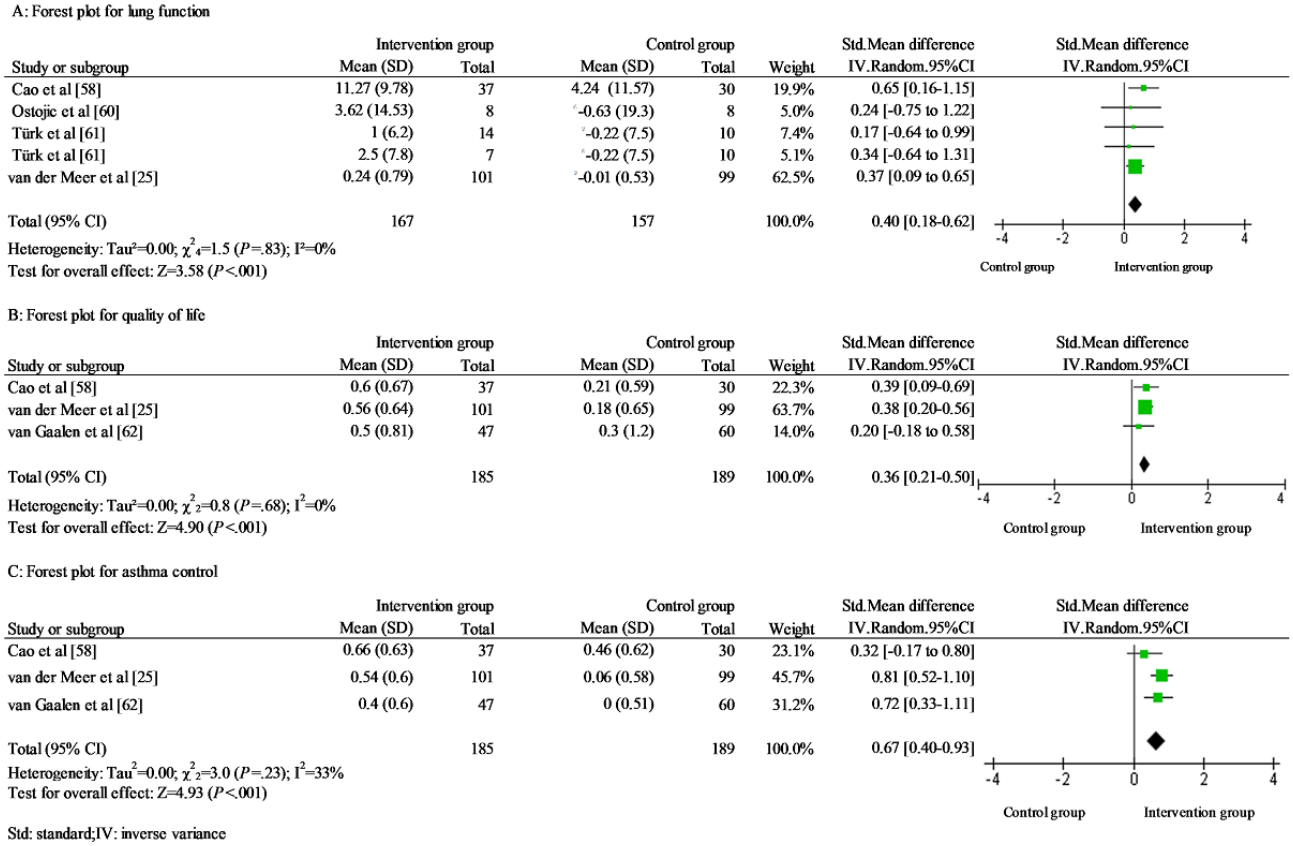


Figure 5. Forest plots for (A) lung function, (B) quality of life, and (C) asthma control in asthma studies.



Discussion

Principal Findings

This systematic review and meta-analysis assessed the effectiveness of blended self-management interventions on health-related effectiveness and process outcome indicators in patients with COPD or asthma. Of the 22 studies that were included in the systematic review, 15 were about COPD and 7 were about asthma.

Studies focusing on COPD patients included 3 different health-related effectiveness outcome indicators, and mixed effects were observed. No effect was observed on mortality. A positive effect was observed for exacerbation frequency and BMI. In total, 11 different process outcome indicators were studied (eg, medication adherence and self-management ability). Of the 3 studies, 2 reported a moderate effect on adherence. A positive effect was found in 1 of the 2 studies on self-management ability. No effects were found on the other process outcomes. Eleven COPD studies were included in the meta-analysis. Blended self-management interventions did not have a significant effect on dyspnea or lung function. Still, they did result in a small improvement in exercise capacity and a moderate improvement in QoL and decreased the admission rate. Overall, the majority of studies had some concerns about the ROB assessment.

The asthma studies included 4 health-related effectiveness outcomes. Large effects were observed in BMI and exercise capacity. There was no effect on the admission rate and exacerbation frequency. Three process outcomes were studied

(ie, visits, intervention and medication adherence, and asthma knowledge). No effect was found on any of the process outcomes. Five asthma studies were included in the meta-analysis. Blended self-management intervention showed a small effect on lung function and QoL, and a moderate effect was found on asthma control. Half of the studies reported some concerns, whereas others showed a high ROB assessment.

The meta-analysis suggested that blended self-management interventions can effectively improve the exercise capacity of patients with COPD. This result was in line with another systematic review that examined the effect of COPD disease management programs, including eHealth and face-to-face components [64]. However, this finding was not consistent with a systematic review of the effect of telehealth in patients with COPD [65]. This may be because the blended programs, contrary to the telehealth programs, were likely to promote exercise capacity using various BCTs, including providing information and instruction on the behavior, self-monitoring, and providing feedback on performance by eHealth and face-to-face intervention [64]. This meta-analysis also showed that blended self-management interventions had a positive effect on QoL, which was in line with the findings of a meta-analysis that investigated the effect of COPD self-management interventions, including various self-management programs [66]. Blended self-management intervention significantly decreased admission rates. This finding was consistent with a previous meta-analysis [67], in which the effect of integrated care from health care providers with or without eHealth was identified. This might be because patients increased their self-management ability and acted on exacerbations more promptly if they received self-management intervention with multiple BCTs [68]. However, the blended self-management interventions included in this meta-analysis did not improve dyspnea and lung function, which was consistent with earlier systematic reviews that investigated the implementation of eHealth or manual therapy in patients with COPD [69,70].

Blended self-management intervention showed an inconsistent impact on process outcomes in patients with COPD. To illustrate, internet-assisted eHealth and individual face-to-face intervention showed a positive effect on self-management ability [54], whereas no effect was found in the blended intervention, including multiple eHealth components and individual face-to-face intervention [51]. The findings in this study may show that certain combinations within the blended interventions may be more effective in some outcomes; however, more large-scale studies using different combinations are needed to provide insight into this. There are several potential explanations for the lack of effects in COPD studies included in the systematic review. First, the length of the blended interventions varied among the included studies (ie, ranged from 4 to 48 weeks). The short intervention duration might have been problematic because patients with mild to very severe COPD were included in the studies. Airway obstruction is usually irreversible in those patients, and the duration of the blended interventions might have been too short to accommodate a change in health [71]. Furthermore, it appears that patients did not adhere sufficiently to blended interventions [22]. This lack of adherence might be because eHealth apps are unfamiliar to some patients [18]. We recommend that future studies educate patients on how to use eHealth because eHealth has a positive effect on improving medication adherence [72].

In asthma studies, in line with other systematic reviews focusing on integrated asthma management (ie, the cooperation of community pharmacists and general practitioners or eHealth and face-to-face intervention), the blended interventions had a positive effect on QoL and asthma control [73,74]. A previous review focusing on face-to-face interventions in patients with asthma showed that face-to-face intervention did not improve QoL and asthma control [75]. The possible reasons for this improvement could be attributed to the integrated care provided by health care providers. Health care providers can update and refer patients for education, counseling, and guidance with eHealth and face-to-face interventions [73,74]. This suggests that, compared with face-to-face interventions, blended interventions or integrated asthma management—where health care providers could refer patients for additional education, counseling, and guidance with eHealth and face-to-face intervention—are more effective. A positive effect was observed on the lung function. This finding was consistent with a meta-analysis that focused on aerobic exercise in patients with asthma [76]. This may be because adequate exercise training is beneficial to lung function. However, due to the limited number of studies included in the meta-analysis, more studies are needed to identify this effect. In this systematic review, limited studies have investigated the effects of blended interventions in patients with asthma. Therefore, the findings should be interpreted cautiously, and future studies with larger sample sizes are needed.

Strengths and Limitations

Several strengths of this review are worth mentioning. First, a detailed description of the interventions was provided, and a wide range of outcomes was included. The detailed information might provide a helpful direction for the development of effective blended self-management interventions. Second, GRADE was used to assess the quality of evidence regarding the true effect of the blended intervention on patients with COPD and asthma. This quality of evidence assessment could provide a clear and pragmatic interpretation of the recommendations for clinicians and policy makers. Finally, we followed a strict study design and precise data analysis steps. By using a strict and precise process, we wanted to ensure the quality of the systematic review and meta-analysis.

However, several limitations also need to be addressed. First, there was a diversity in the intervention and outcome measurements, which made it difficult to compare the findings. Consequently, there may be statistical heterogeneity in the true effect size. Significant heterogeneity potentially diluted the intervention effect [77]. Second, only a small number of studies reported the same outcome measure, and studies with a small sample size were included. These studies may be underpowered to detect a true effect, and this negatively impacted the validity of these studies. Third, the quality of the evidence ranged from very low to high for all outcomes. The various quality of evidence in the outcomes may weaken the recommendation level for clinicians and researchers because the high

heterogeneity among studies downgraded the quality of evidence. Fourth, we were not able to assess the risk of publication bias in the meta-analysis because few studies reported on the same outcome [40]. There may be a potential risk of publication bias. Finally, not all studies reported a follow-up. The lack of this reporting made it impossible to examine the long-term intervention effect in a comprehensive way. The results should be interpreted with caution owing to the limitations mentioned above. Larger RCTs are required to provide more insights, especially RCTs examining the effects of blended interventions in patients with asthma. Moreover, data reporting should be performed in an exact, standardized format to enable reliable extraction for future meta-analysis studies.

Conclusions

The studies focusing on COPD found mixed effects of blended self-management interventions on health-related outcomes, with the strongest evidence found for exercise capacity, QoL, and admission rate. In asthma studies, small to moderate effects were found on asthma control, lung function, and QoL. Overall, blended self-management interventions potentially improve health-related outcomes in patients with COPD and asthma, and more studies are needed to evaluate their effectiveness.

Acknowledgments

This study was funded by the China Scholarship Council.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms.
[DOCX File , 17 KB - jmir_v23i3e24602_app1.docx]

Multimedia Appendix 2
Grading of Recommendations, Assessment, Development, and Evaluation evidence tables.
[DOCX File , 19 KB - jmir_v23i3e24602_app2.docx]

Multimedia Appendix 3
Behavior change techniques in the blended self-management interventions.
[DOCX File , 17 KB - jmir_v23i3e24602_app3.docx]

References

1. Soriano JB, Kendrick PJ, Paulson KR, Gupta V, Vos T, Abrams E. Prevalence and attributable health burden of chronic respiratory diseases, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Respir Med 2020;8(6):585-596.
2. World HO. Chronic obstructive pulmonary disease (COPD). Geneva, Switzerland: WHO 2015;http:2015.
3. Wang HD, Naghavi M, Allen C, Barber RM, Bhutta ZA, Carter A. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Oct 08;388(10053):1459-1544 [FREE Full text] [doi: 10.1016/S0140-6736(16)31012-1] [Medline: 27733281]
4. Mannino DM, Buist AS. Global burden of COPD: risk factors, prevalence, and future trends. The Lancet 2007 Sep;370(9589):765-773. [doi: 10.1016/s0140-6736(07)61380-4]
5. Braman SS. The global burden of asthma. Chest 2006 Jul;130(1 Suppl):4S-12S. [doi: 10.1378/chest.130.1_suppl.4S] [Medline: 16840363]
6. Hosseinzadeh H, Shnaigat M. Effectiveness of chronic obstructive pulmonary disease self-management interventions in primary care settings: a systematic review. Aust. J. Prim. Health 2019;25(3):195-204. [doi: 10.1071/py18181]
7. Raymond B, Luckett T, Johnson M, Hutchinson A, Lovell M, Phillips J. Low-intensity educational interventions supporting self-management to improve outcomes related to chronic breathlessness: a systematic review. npj Prim. Care Respir. Med 2019 Nov 29;29(1):41-49. [doi: 10.1038/s41533-019-0152-8]
8. Lorig KR, Holman HR. Self-management education: History, definition, outcomes, and mechanisms. ann. behav. med 2003 Aug;26(1):1-7. [doi: 10.1207/s15324796abm2601_01]
9. Jonkman NH, Schuurmans MJ, Groenwold RHH, Hoes AW, Trappenburg JCA. Identifying components of self-management interventions that improve health-related quality of life in chronically ill patients: Systematic review and meta-regression analysis. Patient Education and Counseling 2016 Jul;99(7):1087-1098. [doi: 10.1016/j.pec.2016.01.022]

10. Dineen-Griffin S, Garcia-Cardenas V, Williams K, Benrimoj SI. Helping patients help themselves: A systematic review of self-management support strategies in primary health care practice. PLoS One 2019;14(8):e0220116-e0220149 [FREE Full text] [doi: 10.1371/journal.pone.0220116] [Medline: 31369582]
11. Morrison D, Mair FS, Yardley L, Kirby S, Thomas M. Living with asthma and chronic obstructive airways disease: Using technology to support self-management – An overview. Chron Respir Dis 2016 Aug 10;14(4):407-419. [doi: 10.1177/1479972316660977]
12. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun 18;3(2):e20-e22. [doi: 10.2196/jmir.3.2.e20]
13. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van RL, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014;16(4):e110-e133 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]
14. Morrison D, Wyke S, Agur K, Cameron EJ, Docking RI, MacKenzie AM, et al. Digital Asthma Self-Management Interventions: A Systematic Review. J Med Internet Res 2014 Feb 18;16(2):e51-e71. [doi: 10.2196/jmir.2814]
15. Talboom-Kamp E, Holstege MS, Chavannes NH, Kasteleyn MJ. Effects of use of an eHealth platform e-Vita for COPD patients on disease specific quality of life domains. Respir Res 2019 Jul 10;20(1):146-155. [doi: 10.1186/s12931-019-1110-2]
16. Hallensleben C, van Luenen S, Rolink E, Ossebaard HC, Chavannes NH. eHealth for people with COPD in the Netherlands: a scoping review. COPD 2019 Jul;Volume 14:1681-1690. [doi: 10.2147/copd.s207187]
17. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using Internet and Mobile Phone Technology to Deliver an Automated Physical Activity Program: Randomized Controlled Trial. J Med Internet Res 2007 Apr 27;9(2):e7-31. [doi: 10.2196/jmir.9.2.e7]
18. Ariens LFM, Schussler-Raymakers FML, Frima C, Flinterman A, Hamminga E, Arents BWM, et al. Barriers and Facilitators to eHealth Use in Daily Practice: Perspectives of Patients and Professionals in Dermatology. J Med Internet Res 2017 Sep 05;19(9):e300. [doi: 10.2196/jmir.7512]
19. Kloek C, Bossen D, de Bakker DH, Veenhof C, Dekker J. Blended Interventions to Change Behavior in Patients With Chronic Somatic Disorders: Systematic Review. J Med Internet Res 2017 Dec 21;19(12):e418 [FREE Full text] [doi: 10.2196/jmir.8108] [Medline: 29269338]
20. Erbe D, Eichert HC, Riper H, Ebert DD. Blending Face-to-Face and Internet-Based Interventions for the Treatment of Mental Disorders in Adults: Systematic Review. J Med Internet Res 2017 Sep 15;19(9):e306. [doi: 10.2196/jmir.6588]
21. Rasing SPA, Stikkelbroek YAJ, Bodden DHM. Is Digital Treatment the Holy Grail? Literature Review on Computerized and Blended Treatment for Depressive Disorders in Youth. Int J Environ Res Public Health 2019 Dec 24;17(1):e306 [FREE Full text] [doi: 10.3390/ijerph17010153] [Medline: 31878249]
22. Garcia-Aymerich J, Hernandez C, Alonso A, Casas A, Rodriguez-Roisin R, Anto JM, et al. Effects of an integrated care intervention on risk factors of COPD readmission. Respir Med 2007 Jul;101(7):1462-1469 [FREE Full text] [doi: 10.1016/j.rmed.2007.01.012] [Medline: 17339106]
23. Barbanel D, Eldridge S, Griffiths C. Can a self-management programme delivered by a community pharmacist improve asthma control? A randomised trial. Thorax 2003 Oct;58(10):851-854 [FREE Full text] [doi: 10.1136/thorax.58.10.851] [Medline: 14514935]
24. Koff PB, Jones RH, Cashman JM, Voelkel NF, Vandivier RW. Proactive integrated care improves quality of life in patients with COPD. European Respiratory Journal 2009 Jan 07;33(5):1031-1038. [doi: 10.1183/09031936.00063108]
25. van der Meer V, Bakker MJ, van den Hout WB, Rabe KF, Sterk PJ, Kievit J, et al. Internet-based self-management plus education compared with usual care in asthma: a randomized trial. Ann Intern Med 2009 Jul 21;151(2):110-120. [Medline: 19620163]
26. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. International Journal of Surgery 2010;8(5):336-341. [doi: 10.1016/j.ijsu.2010.02.007]
27. Frandsen TF, Nielsen MFB, Lindhardt CL, Eriksen MB. Using the full PICO model as a search tool for systematic reviews resulted in lower recall for some PICO elements. Journal of Clinical Epidemiology 2020 Nov;127:69-75. [doi: 10.1016/j.jclinepi.2020.07.005]
28. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012:276-282. [doi: 10.11613/bm.2012.031]
29. Mirza S, Clay RD, Koslow MA, Scanlon PD. COPD Guidelines: A Review of the 2018 GOLD Report. Mayo Clinic Proceedings 2018 Oct;93(10):1488-1502. [doi: 10.1016/j.mayocp.2018.05.026]
30. Chaet AV, Morshedi B, Wells KJ, Barnes LE, Valdez R. Spanish-Language Consumer Health Information Technology Interventions: A Systematic Review. J Med Internet Res 2016 Dec 10;18(8):e214 [FREE Full text] [doi: 10.2196/jmir.5794] [Medline: 27511437]
31. Gillick MR. The critical role of caregivers in achieving patient-centered care. JAMA 2013 Aug 14;310(6):575-576. [doi: 10.1001/jama.2013.7310] [Medline: 23867885]
32. Noordman J, van der Weijden T, van Dulmen S. Communication-related behavior change techniques used in face-to-face lifestyle interventions in primary care: A systematic review of the literature. Patient Education and Counseling 2012 Nov;89(2):227-244. [doi: 10.1016/j.pec.2012.07.006]
33. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28:l4898. [doi: 10.1136/bmj.l4898]

34. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Schünemann HJ. What is "quality of evidence" and why is it important to clinicians? BMJ 2008 May 01;336(7651):995-998. [doi: 10.1136/bmj.39490.551019.be]
35. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. Journal of Clinical Epidemiology 2011 Apr;64(4):401-406. [doi: 10.1016/j.jclinepi.2010.07.015]
36. Crutzen R. Adding effect sizes to a systematic review on interventions for promoting physical activity among European teenagers. Int J Behav Nutr Phys Act 2010;7(1):29. [doi: 10.1186/1479-5868-7-29]
37. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [Medline: 19565683]
38. Lang TA, Lang T, Secic M. How to Report Statistics in Medicine. The Nurse Practitioner 1997;22(5):198. [doi: 10.1097/00006205-199705000-00022]
39. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can Smartphone Apps Increase Physical Activity? Systematic Review and Meta-Analysis. J Med Internet Res 2019 Mar 19;21(3):e12053. [doi: 10.2196/12053]
40. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [Medline: 9310563]
41. DerSimonian R, Laird N. Meta-analysis in clinical trials revisited. Contemp Clin Trials 2015 Nov;45(Pt A):139-145 [FREE Full text] [doi: 10.1016/j.cct.2015.09.002] [Medline: 26343745]
42. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Cochrane. 2018. URL: https://training.cochrane.org/search/site/cochrane%20handbook%20systematic%20reviews%20interventions [accessed 2020-03-09]
43. Viechtbauer W, Cheung MW. Outlier and influence diagnostics for meta-analysis. Res Synth Methods 2010 Apr;1(2):112-125. [doi: 10.1002/jrsm.11] [Medline: 26061377]
44. Wallace BC, Schmid CH, Lau J, Trikalinos TA. Meta-Analyst: software for meta-analysis of binary, continuous and diagnostic data. BMC Med Res Methodol 2009 Dec 4;9(1):80-91. [doi: 10.1186/1471-2288-9-80]
45. Bentley CL, Mountain GA, Thompson J, Fitzsimmons DA, Lowrie K, Parker SG, et al. A pilot randomised controlled trial of a Telehealth intervention in patients with chronic obstructive pulmonary disease: challenges of clinician-led data collection. Trials 2014;15:313 [FREE Full text] [doi: 10.1186/1745-6215-15-313] [Medline: 25100550]
46. Cameron-Tucker HL, Wood-Baker R, Joseph L, Walters JA, Schüz N, Walters EH. A randomized controlled trial of telephone-mentoring with home-based walking preceding rehabilitation in COPD. Int J Chron Obstruct Pulmon Dis 2016;11:1991-2000 [FREE Full text] [doi: 10.2147/COPD.S109820] [Medline: 27601892]
47. Casas A, Troosters T, Garcia-Aymerich J, Roca J, Hernández C, Alonso A, et al. Integrated care prevents hospitalisations for exacerbations in COPD patients. Eur Respir J 2006 Jul;28(1):123-130 [FREE Full text] [doi: 10.1183/09031936.06.00063205] [Medline: 16611656]
48. Chau JPC, Lee DTF, Yu DSF, Chow AYM, Yu WC, Chair SY, et al. A feasibility study to investigate the acceptability and potential effectiveness of a telecare service for older people with chronic obstructive pulmonary disease. International Journal of Medical Informatics 2012 Oct;81(10):674-682. [doi: 10.1016/j.ijmedinf.2012.06.003]
49. Hæsum LKE, Ehlers LH, Hejlesen OK. The long-term effects of using telehomecare technology on functional health literacy: results from a randomized trial. Public Health 2017 Sep;150:43-50. [doi: 10.1016/j.puhe.2017.05.002] [Medline: 28623766]
50. Jehn M, Donaldson G, Kiran B, Liebers U, Mueller K, Scherer D, et al. Tele-monitoring reduces exacerbation of COPD in the context of climate change--a randomized controlled trial. Environ Health 2013;12:99 [FREE Full text] [doi: 10.1186/1476-069X-12-99] [Medline: 24261700]
51. Nguyen HQ, Donesky-Cuenco D, Wolpin S, Reinke LF, Benditt JO, Paul SM, et al. Randomized controlled trial of an internet-based versus face-to-face dyspnea self-management program for patients with chronic obstructive pulmonary disease: pilot study. J Med Internet Res 2008;10(2):e9 [FREE Full text] [doi: 10.2196/jmir.990] [Medline: 18417444]
52. Sorknaes AD, Bech M, Madsen H, Titlestad IL, Hounsgaard L, Hansen-Nord M, et al. The effect of real-time teleconsultations between hospital-based nurses and patients with severe COPD discharged after an exacerbation. J Telemed Telecare 2013 Dec;19(8):466-474. [doi: 10.1177/1357633X13512067] [Medline: 24227799]
53. Stamenova V, Liang K, Yang R, Engel K, van Lieshout F, Lalingo E. Technology-Enabled Self-Management of Chronic Obstructive Pulmonary Disease With or Without Asynchronous Remote Monitoring: Randomized Controlled Trial. J Med Internet Res, 2020 2020;22(7):e18598. [doi: 10.2196/preprints.18598]
54. Wang L, He L, Tao Y, Sun L, Zheng H, Zheng Y, et al. Evaluating a Web-Based Coaching Program Using Electronic Health Records for Patients With Chronic Obstructive Pulmonary Disease in China: Randomized Controlled Trial. J Med Internet Res 2017 Jul 21;19(7):e264 [FREE Full text] [doi: 10.2196/jmir.6743] [Medline: 28733270]
55. Wang LH, Guo YM, Wang ML, Zhao Y. A mobile health application to support self-management in patients with chronic obstructive pulmonary disease: a randomised controlled trial. Clin Rehabil 2020 Sep 09;35(1):90-101. [doi: 10.1177/0269215520946931]
56. Wei L, Yang XY, Li J, Liu LH, Luo HY, Zheng ZG. Effect of pharmaceutical care on medication adherence and hospital admission in patients with chronic obstructive pulmonary disease (COPD): a randomized controlled study. J Thorac Dis 2014;6(6):656-662.

57. Xin CW, Xia ZN, Jiang C, Lin MM, Li GH. The impact of pharmacist-managed clinic on medication adherence and health-related quality of life in patients with COPD: a randomized controlled study. PPA 2016 Jul;Volume 10:1197-1203. [doi: 10.2147/ppa.s110167]
58. Cao Y, Lin SH, Zhu D, Xu F, Chen ZH, Shen HH, et al. WeChat Public Account Use Improves Clinical Control of Cough-Variant Asthma: A Randomized Controlled Trial. Med Sci Monit 2018 Mar 14;24:1524-1532 [FREE Full text] [doi: 10.12659/msm.907284] [Medline: 29536984]
59. Kohler B, Kellerer C, Schultz K, Wittmann M, Atmann O, Linde K. An Internet-Based Asthma Self-Management Program Increases Knowledge About Asthma Results of a Randomized Controlled Trial. Dtsch Arztebl Int 2020;117(5):64-71. [doi: 10.3238/arztebl.2020.0064]
60. Ostojic V, Cvoriscec B, Ostojic SB, Reznikoff D, Stipic-Markovic A, Tudjman Z. Improving asthma control through telemedicine: a study of short-message service. Telemed J E Health 2005 Feb;11(1):28-35. [doi: 10.1089/tmj.2005.11.28] [Medline: 15785218]
61. Türk Y, Theel W, van Huisstede A, van de Geijn GM, Birnie E, Hiemstra PS, et al. Short-term and long-term effect of a high-intensity pulmonary rehabilitation programme in obese patients with asthma: a randomised controlled trial. Eur Respir J 2020 Jul;56(1):1901820. [doi: 10.1183/13993003.01820-2019] [Medline: 32299852]
62. van Gaalen JL, Beerthuizen T, van der Meer MV, van Reisen P, Redelijkheid GW, Snoeck-Stroband JB, et al. Long-term outcomes of internet-based self-management support in adults with asthma: randomized controlled trial. J Med Internet Res 2013;15(9):e188 [FREE Full text] [doi: 10.2196/jmir.2640] [Medline: 24028826]
63. Sundberg R, Tunsäter A, Palmqvist M, Ellbjär S, Löwhagen O, Torén K. A randomized controlled study of a computerized limited education program among young adults with asthma. Respir Med 2005 Mar;99(3):321-328 [FREE Full text] [doi: 10.1016/j.rmed.2004.08.006] [Medline: 15733508]
64. Peytremann-Bridevaux I, Staeger P, Bridevaux PO, Ghali WA, Burnand B. Effectiveness of Chronic Obstructive Pulmonary Disease-Management Programs: Systematic Review and Meta-Analysis. The American Journal of Medicine 2008 May;121(5):433-443.e4. [doi: 10.1016/j.amjmed.2008.02.009]
65. Lundell S, Holmner A, Rehn B, Nyberg A, Wadell K. Telehealthcare in COPD: a systematic review and meta-analysis on physical outcomes and dyspnea. Respir Med 2015 Jan;109(1):11-26. [doi: 10.1016/j.rmed.2014.10.008] [Medline: 25464906]
66. Cannon D, Buys N, Sriram KB, Sharma S, Morris N, Sun J. The effects of chronic obstructive pulmonary disease self-management interventions on improvement of quality of life in COPD patients: A meta-analysis. Respir Med 2016 Dec;121:81-90 [FREE Full text] [doi: 10.1016/j.rmed.2016.11.005] [Medline: 27888996]
67. Long H, Howells K, Peters S, Blakemore A. Does health coaching improve health-related quality of life and reduce hospital admissions in people with chronic obstructive pulmonary disease? A systematic review and meta-analysis. Br J Health Psychol 2019 Sep;24(3):515-546 [FREE Full text] [doi: 10.1111/bjhp.12366] [Medline: 31033121]
68. Ridwan ES, Hadi H, Wu YL, Tsai PS. Effects of Transitional Care on Hospital Readmission and Mortality Rate in Subjects With COPD: A Systematic Review and Meta-Analysis. Respir Care 2019 Aug 29;64(9):1146-1156. [doi: 10.4187/respcare.06959]
69. Simonelli C, Vitacca M, Vignoni M, Ambrosino N, Paneroni M. Effectiveness of manual therapy in COPD: A systematic review of randomised controlled trials. Pulmonology 2019 Jul;25(4):236-247. [doi: 10.1016/j.pulmoe.2018.12.008]
70. Sul AR, Lyu DH, Park DA. Effectiveness of telemonitoring versus usual care for chronic obstructive pulmonary disease: A systematic review and meta-analysis. J Telemed Telecare 2018 Dec 12;26(4):189-199. [doi: 10.1177/1357633x18811757]
71. Imamura S, Inagaki T, Terada J, Nagashima K, Katsura H, Tatsumi K. Long-term efficacy of pulmonary rehabilitation with home-based or low frequent maintenance programs in patients with chronic obstructive pulmonary disease: a meta-analysis. Ann Palliat Med 2020 Sep;9(5):2606-2615. [doi: 10.21037/apm-19-581]
72. Murphie P, Little S, McKinstry B, Pinnock H. Remote consulting with telemonitoring of continuous positive airway pressure usage data for the routine review of people with obstructive sleep apnoea hypopnoea syndrome: A systematic review. J Telemed Telecare 2019 Jan;25(1):17-25. [doi: 10.1177/1357633X17735618] [Medline: 28990455]
73. Mubarak N, Hatah E, Khan TM, Zin CS. A systematic review and meta-analysis of the impact of collaborative practice between community pharmacist and general practitioner on asthma management. J Asthma Allergy 2019;12:109-153 [FREE Full text] [doi: 10.2147/JAA.S202183] [Medline: 31213852]
74. Chongmelaxme B, Lee S, Dhippayom T, Saokaew S, Chaiyakunapruk N, Dilokthornsakul P. The Effects of Telemedicine on Asthma Control and Patients' Quality of Life in Adults: A Systematic Review and Meta-analysis. The Journal of Allergy and Clinical Immunology: In Practice 2019 Jan;7(1):199-216.e11. [doi: 10.1016/j.jaip.2018.07.015]
75. Baishnab E, Karner C. Primary care based clinics for asthma. Cochrane Database Syst Rev 2012. [doi: 10.1002/14651858.cd003533.pub2]
76. Hansen ESH, Pitzner-Fabricius A, Toennesen LL, Rasmusen HK, Hostrup M, Hellsten Y, et al. Effect of aerobic exercise training on asthma in adults: a systematic review and meta-analysis. Eur Respir J 2020 Apr 29;56(1):2000146. [doi: 10.1183/13993003.00146-2020]
77. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. Journal of Clinical Epidemiology 2011 Dec;64(12):1294-1302. [doi: 10.1016/j.jclinepi.2011.03.017]

Abbreviations

BCT: behavior change technique
CAT: chronic obstructive pulmonary disease assessment test
CG: control group
CONSORT: Consolidated Standards of Reporting Trials
COPD: chronic obstructive pulmonary disease
CRQ: chronic respiratory questionnaire
DALY: disability-adjusted life year
FEV1: forced expiratory volume in 1s
FVC: forced vital capacity
GOLD: Global Initiative for Chronic Obstructive Lung Disease
GRADE: Grading of Recommendations, Assessment, Development, and Evaluation
IG: intervention group
PC: primary care
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QALY: quality-adjusted life year
QoL: quality of life
RCT: randomized controlled trial
ROB: risk of bias
RR: relative ratio
SC: secondary care
SGRQ: Saint-George's Respiratory Questionnaire
SMD: standardized mean difference
UC: usual care

Edited by G Eysenbach; submitted 26.09.20; peer-reviewed by T Ndabu, Y Wang, D Erbe, R Bajpai, E Drokow; comments to author 26.11.20; revised version received 17.01.21; accepted 18.01.21; published 31.03.21.

Please cite as:
Song X, Hallensleben C, Zhang W, Jiang Z, Shen H, Gobbens RJJ, Kleij RMJJVD, Chavannes NH, Versluis A
Blended Self-Management Interventions to Reduce Disease Burden in Patients With Chronic Obstructive Pulmonary Disease and Asthma: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(3):e24602
URL: https://www.jmir.org/2021/3/e24602
doi:10.2196/24602
PMID:33788700

Viewpoint

Guidelines for Conducting Virtual Cognitive Interviews During a Pandemic

James A Shepperd[1], PhD; Gabrielle Pogge[1], PhD; Jean M Hunleth[2], MPH; Sienna Ruiz[2], BA; Erika A Waters[2], PhD

[1]Department of Psychology, University of Florida, Gainesville, FL, United States

[2]Department of Surgery, Division of Public Health Sciences, Washington University, St. Louis, MO, United States

Corresponding Author:
James A Shepperd, PhD
Department of Psychology
University of Florida
945 Center Drive
Gainesville, FL, 32611
United States
Phone: 1 352 273 2165
Email: shepperd@ufl.edu

Abstract

The COVID-19 pandemic has challenged researchers working in physical contact with research participants. Cognitive interviews examine whether study components (most often questionnaire items) are worded or structured in a manner that allows study participants to interpret the items in a way intended by the researcher. We developed guidelines to conduct cognitive interviews virtually to accommodate interviewees who have limited access to the internet. The guidelines describe the essential communication and safety equipment requirements and outline a procedure for collecting responses while maintaining the safety of the participants and researchers. Furthermore, the guidelines provide suggestions regarding training of participants to use the technology, encouraging them to respond aloud (a potential challenge given that the researcher is not physically present with the participant), and testing and deploying the equipment prior to the interview. Finally, the guidelines emphasize the need to adapt the interview to the circumstances and anticipate potential problems that might arise.

(J Med Internet Res 2021;23(3):e25173) doi:10.2196/25173

KEYWORDS

cognitive interview; COVID-19; guidelines; teleresearch; pandemic; tablet computer; telehealth; training

Introduction

Infectious disease pandemics can potentially derail studies involving in-person interactions with participants, such as cognitive interviews. Cognitive interviews examine whether study components—most often questionnaire items—are worded or structured in a manner that allows study participants to interpret the items in a way intended by the researcher [1]. The interviews allow researchers to fix problems before fielding the survey. Cognitive interviews often involve a "think aloud" protocol where researchers ask participants to think aloud as they read items and reason through their responses. Although the think aloud protocol is challenging and may not be feasible for some participants, it provides insights into how participants interpret items. Researchers can pursue potentially concerning responses with additional verbal questions to identify the point of confusion, and they can explore possible alternative items or instruction wording. Cognitive interviews can also help determine whether certain words, concepts, or phrases are understood similarly across participants, whether a potentially sensitive item is offensive to participants, and whether items require adaptation to accommodate individuals with limited literacy or health literacy.

COVID-19 was declared a national emergency in the United States in March 2020, prompting various social distancing protocols and other restrictions on in-person contact, including a moratorium on in-person human subject research imposed by institutional review boards (IRBs). Videoconferencing platforms including Zoom (Zoom Video Communications), GoToMeeting (LogMeIn Inc), Webex (Cisco Webex), and Skype (Skype Technologies, Microsoft Corp) offer temporary solutions to some researchers in that they facilitate web-based interactions with study participants. However, these platforms require that participants have a web-accessible device, reliable and sufficiently high-speed internet access to support videoconferencing platforms, and a registered account on such

platforms and software to access these platforms. Participants with low income and participants who reside at remote locations often do not meet these requirements [2].

At our institutions, COVID-19 led to temporary suspension of all research activities involving human subjects just as we began in-person cognitive interviews for two nationwide surveys. Our participants were caregivers of children with asthma, and many of the caregivers had asthma themselves or other health problems that made in-person interviews hazardous for them. Some of our study staff also had health conditions that placed them at an increased risk of COVID-19–related mortality. In accordance with the IRB restrictions, and to ensure the safety of the participants, their families, and our staff and their families, we developed a minimal contact protocol. The protocol entailed conducting a virtual interview, which can have more advantages than in-person interviews: virtual interviews can be more convenient for researchers and participants; lead to the inclusion of participants who might otherwise be excluded, such as people with disabilities or people who live in remote areas [3]; yield data similar in quality to data obtained through in-person interviews [4]; and—because of the perceived anonymity—facilitate discussions on sensitive topics [5].

We initially developed a protocol to conduct cognitive interviews with caregivers. However, we have since expanded the protocol to interview children with asthma. The challenges researchers encounter undoubtedly vary among studies. We describe some general guidelines that we developed to facilitate successful remote interviews.

Overview of the Protocol

The cognitive interviews proceeded as follows. Our scheduler called potential participants, invited them to participate in the interview, and then scheduled an interview. The researcher then called the participants the day before the interview to provide them with more details regarding the procedures and safety protocol, and to establish a rapport. The researcher arrived at the participant's residence and, while still seated in the car, set up the equipment (a laptop, tablet computer, and a portable hotspot device), launched the meeting, and ensured that the tablet displayed the survey and that the tablet screen was visible on all devices involved in the meeting. A research assistant joined the meeting virtually during the setup. Next, the researcher delivered the tablet and portable hotspot device to the participant and returned to the car (or another socially distanced location) to conduct the interview. After the interview, the researcher retrieved the equipment and concluded the meeting. Our sample comprised 8 caregivers (7 women, 1 man; 5 Black individuals, 3 White individuals; aged 33-49 years) of children with asthma and living in Gainesville, FL. All caregivers had low income, but none had impairments that affected their ability to use the equipment. All caregivers had participated in a prior at-home interview with the members of our research group (albeit not the current researcher). None of them declined to participate.

The Guidelines

Equipment

We needed two types of equipment: web-enabled communication equipment and safety equipment to prevent the spread of COVID-19. Specifically, we required the following:

1. A survey platform: we used Qualtrics XM (Qualtrics), although SurveyMonkey (SVMK Inc) is equally suitable and free of charge.
2. A portable hotspot device that provides internet access to wireless devices: we purchased a Verizon Jetpack MiFi8800L device (Verizon Wireless) with a monthly contract of US $35.
3. A videoconferencing platform: we used Zoom.
4. A laptop to launch the conference call and to communicate with participants during the interview.
5. Headphones for the laptop to reduce background noise: we used a microphone headset that was previously obtained with an iPhone (Apple Inc) but costs US $11 when purchased separately.
6. A tablet computer for participants to access the internet: we purchased the Apple iPad Air 3rd generation (64 GB, 10.5 inch, Wi-Fi) for US $479 and added a screen protector for US $11, a replacement warranty for US $59, and a protective case (Seymac Co Ltd) for US $19. The screen protector, warranty, and protective case were essential because we could not risk losing the tablet because of accidental damage.
7. Safety equipment: we purchased protective masks, disinfectant wipes, and hand sanitizers, and we placed pens, payment forms, and payment cards (ie, debit cards) in a zip-lock bag.

The total communication equipment cost was US $603 and the total safety equipment cost was US $15.

Participant Preparation

Most participants had limited or no experience with using the portable hotspot device, tablet computer, web-based videoconferencing platform, or with completing a web-based survey. To ensure a smooth flow of the interview, we called participants the day before the interview and briefed them on the procedures and the safety protocol. We informed them of the number and types of items, noted when the session would begin, how long it would take, how we would compensate them for their participation, and how the tablet computer, hotspot device, survey platform, and videoconferencing platform operated. We also explained our COVID-19 safety protocol: we would wear masks and maintain a distance of 6 feet during interpersonal interaction, and we would use disinfectant wipes to clean the iPad and portable hotspot device before delivering them to and after retrieving them from participants. We acknowledged that although wearing masks and maintaining social distance might feel awkward and uncomfortable, the university required that we adhere to these steps to prevent COVID-19 transmission.

This advance phone call provided us an opportunity to build a rapport with the participants. The researcher underscored the

value of the participants' contribution and emphasized our desire to learn from their expertise as caregivers and our need to receive feedback on how to best ask our survey questions. The researcher endeavored to bond with participants by establishing a friendly tone, a sense of comradery, a shared goal of addressing a health concern, and an understanding that the participants' views were critical and made a difference (the researcher offered examples of items that were changed based on participant feedback).

Equipment Preparation

The researcher called participants 1 hour prior to the interview to remind them about the interview, and then arrived at the participants' residence at least 10-15 minutes early to ensure adequate time to set up the devices and the videoconferencing platform, and to log into the meeting on the laptop and tablet computer. Setting up the equipment required several steps:

1. Turn on the portable hotspot device, tablet computer, and laptop, and ensure that the laptop and tablet computer both access the internet through the portable hotspot device and not from some other wireless device.
2. Open the survey link on the tablet computer. Weblinks to the survey and the web-based meeting can be lengthy. It is often easy to email the links to the tablet computer and then click the links obtained from the email. However, one must take care to log out of the email account before delivering the tablet to participants to prevent participants from accessing the email account through the device. Furthermore, the researcher should also disable any alerts on the tablet computer so that participants are not interrupted during the interview. Instructions for disabling or hiding email accounts on tablet computers are available online.
3. Launch the meeting on the laptop (which allows the researcher to control screen sharing during the meeting) and join the meeting from the tablet computer. Only the researcher and research assistant had access to the URL for the Zoom meeting, which was typically generated 1 hour before the meeting. Although we did not password-protect the meetings, researchers concerned with privacy invasion can do so. In addition, the network connection provided by the portable hotspot device was password protected. Finally, we collected no participant-identifiable information in the survey; Qualtrics encrypts responses using secure socket layers and masks all IP addresses, thus providing the researchers access to only the survey responses.
4. Allow screen sharing from the laptop, and then share the screen for the survey link from the tablet computer. The screen sharing allows the researchers to monitor participants' responses to the survey in real time and probe them as necessary.
5. Test the audio in the meeting. An audio test can be challenging because of the possibility of generating a feedback loop when the laptop and tablet computer are both logged into the virtual meeting and are proximal to each other. If the sound works properly for both the tablet and the laptop, the researcher can deliver the tablet computer and portable hotspot device to the participant.

The researcher performs these tasks in the car, which requires some juggling. We found it useful to practice setting up the devices in the car at home before proceeding to the participants' residence. The research assistant took notes and asked participants additional questions if needed. Because seeing the researchers' faces while taking the survey could be distracting and may affect the participant responses, both researchers disabled their cameras throughout the meeting.

Adapting the Interview to the Circumstances

Once the researcher completed the steps successfully, the researcher delivered the tablet computer and hotspot device to the participant. The researcher did not enter the participant's residence, but rather stayed outside and cleaned the devices with a disinfectant wipe in front of the participant. If possible, the researcher placed the devices on a porch table or another surface and stepped back rather than handing them directly to the participant. The researcher then introduced the research assistant (who was audible through the tablet) to the participant and answered any questions. This point in time was an opportunity to review the safety protocol with the participant.

After explaining the safety protocol and responding to the participant's questions, the researcher moved to a distant site (often returning to the car) to proceed with the interview. During this brief transition, the research assistant, speaking through the tablet, reminded the participant of the procedure and reiterated the value of their participation. Many apartment complexes where we conducted the interviews had picnic tables where the researcher could conduct the interviews while the participant completed the survey in their residence. However, sitting outside underscored the need for a microphone headset. In several instances, the researcher was interrupted by other people (eg, the apartment manager, other residents who were being social) while sitting at an outdoor picnic table. Moreover, some apartment complexes were noisy with barking dogs, neighbors talking, and street sounds. Without the microphone headset, these interruptions and noises would be distracting to the researcher and to the participant who can hear through the tablet computer what the researcher hears.

We had only the survey displayed on the tablet computer screen so that we could monitor participants' responses. We asked participants to read each item on the survey aloud, verbally declare their response, and explain aloud the reason for their response. Reminders were sometimes necessary, yet participants acclimated rapidly to this request even though the researcher was not physically present with them. Each page typically contained 2-10 items, and we stopped participants at the end of each page to probe their responses on the page in more detail and to ask what certain phrases or words meant to them. Having participants talk aloud shortened the interview durations because in many instances, participants had already explained their responses, eliminating our need for further probing.

Once the survey was complete, the researcher retrieved the tablet computer and hotspot device and provided the participant with a zip-lock bag containing a debit card, pen, and payment receipt form. The participant used the pen to sign the payment form, returned the pen and form in the bag, and retained the debit card. As a final gesture, and because the participant

handled the zip-lock bag, the researcher offered a squirt of hand sanitizer before leaving. The researcher then returned to the car, sanitized the equipment and their hands, turned off the hotspot device, closed the browsers for the survey and the meeting on the tablet computer, and ended the meeting on the laptop.

Managing Potential Problems

Unplanned events are inevitable, and the researcher must be prepared to troubleshoot. On two occasions, the portable hotspot device failed: on one occasion, an electrical storm disrupted service for a few seconds, and on another occasion, the portable hotspot device overheated from sustained exposure to the hot sun. In the latter instance, it took a couple of minutes before the hotspot device cooled down and resumed functioning. Occasionally, participants were not wearing masks or wearing them around their chin or neck and the interviewer reminded them to wear a mask or to wear it correctly. We encountered no resistance regarding the safety protocol, perhaps because we were clear in the phone conversations that the university required us to follow the safety protocol, that it was for all our benefit, and that we were all obliged to follow it. In all instances, it was clear that the participant merely forgot to follow the safety protocol.

Portable hotspot devices have limited broadband capacity, and videoconferencing draws considerable bandwidth. If the researcher and participant are both accessing the internet through the hotspot device, they are more likely to experience disrupted internet access. However, these disruptions did not occur for us if only the participant used the video mode and if the researcher closed all other web-based programs on the laptop (such as email platforms). It is noteworthy that Wi-Fi speed is generally low for everyone if too many people in a location attempt to use it simultaneously. Finally, participants sometimes clicked on a button on the tablet computer, which directed them away from the survey, or the tablet computer entered sleep mode during extended periods of conversation (although the audio was still retained throughout the meeting). We addressed all problems rapidly by instructing the participant how to return to the survey. If the researcher had to briefly retrieve the tablet computer or hotspot device, he/she had extra disinfectant wipes for cleaning the surfaces.

We considered having the participant's face displayed on the screen while they completed the interview, thereby allowing us to monitor their nonverbal responses. Attending to a participant's nonverbal responses is a vital component of cognitive interviews. Nonverbal cues can reveal confusion (eg, a furrowed brow) and boredom or discomfort (eg, exaggerated sighing). However, we found it challenging enough to monitor their responses to the items, and the video quality was insufficient for monitoring and interpreting facial expressions. Thus, we instead attended closely to participants' questionnaire responses and their verbalizations. Modulations and inflections in participants' voices revealed a wealth of useful information (eg, surprise, confusion, incredulity, or annoyance) independent of the content. In addition, participants occasionally took more time to respond to certain survey items, which suggested that they were perhaps struggling with those items. These instances prompted us to ask participants to share with us what slowed them down.

Finally, we were concerned that the participants might be less responsive to researchers conversing with them in an unfamiliar format. However, we encountered no such problems, presumably because of the efforts we took to establish a rapport with participants and because other members of our team had interviewed these participants in the past and thus had established a relationship. We also speculate that participating from one's own home was comforting, and conversing with remote researchers who could be heard but not be seen generated a sense of privacy and intimacy that fostered greater disclosure.

Summary

The COVID-19 pandemic has posed challenges among researchers conducting cognitive interviews, particularly in populations with limited access to the internet, an internet accessible device, or web-based videoconferencing platforms. Our guidelines describe how researchers can address these challenges and continue performing studies involving cognitive interviews. These guidelines describe some necessary communication and safety equipment and outline a procedure for collecting responses while maintaining the safety of the participant and researcher. These guidelines also provide tips for establishing rapport, training participants in the technology, encouraging participants to respond aloud, and testing and deploying the equipment prior to an interview. Finally, the guidelines emphasize the need to adapt the interview to various circumstances and anticipate potential unplanned events.

Acknowledgments

This study was funded by the National Heart, Lung, and Blood Institute of the National Institutes of Health (R01HL137680; MPI: JAS and EAW).

Conflicts of Interest

None declared.

References

1. Willis G. Cognitive Interviewing. Thousand Oaks, CA: Sage Publications; 2005.
2. Anderson M. Mobile Technology and Home Broadband 2019. Pew Research Center. 2019 Jun 13. URL: https://www.pewresearch.org/internet/2019/06/13/mobile-technology-and-home-broadband-2019/ [accessed 2021-02-17]

3. Wilkerson JM, Iantaffi A, Grey JA, Bockting WO, Rosser BRS. Recommendations for internet-based qualitative health research with hard-to-reach populations. Qual Health Res 2014 Apr;24(4):561-574 [FREE Full text] [doi: 10.1177/1049732314524635] [Medline: 24623662]
4. Shapka JD, Domene JF, Khan S, Yang LM. Online versus in-person interviews with adolescents: An exploration of data equivalence. Computers in Human Behavior 2016 May;58:361-367. [doi: 10.1016/j.chb.2016.01.016]
5. O'Reilly M, Dogra N. Interviewing Children and Young People for Research. London, UK: Sage Publications; 2017.

Abbreviations

IRB: institutional review board

Edited by G Eysenbach; submitted 20.10.20; peer-reviewed by T Aslanidis, D Cooke, H Henshaw; comments to author 18.11.20; revised version received 25.11.20; accepted 30.11.20; published 11.03.21.

Please cite as:
Shepperd JA, Pogge G, Hunleth JM, Ruiz S, Waters EA
Guidelines for Conducting Virtual Cognitive Interviews During a Pandemic
J Med Internet Res 2021;23(3):e25173
URL: https://www.jmir.org/2021/3/e25173
doi:10.2196/25173
PMID:33577464

Viewpoint

HOPES: An Integrative Digital Phenotyping Platform for Data Collection, Monitoring, and Machine Learning

Xuancong Wang[1], PhD; Nikola Vouk[1], MBA; Creighton Heaukulani[1], PhD; Thisum Buddhika[1], MComp; Wijaya Martanto[1], MD, PhD; Jimmy Lee[2,3], MBBS, MMed, MCI; Robert JT Morris[1,4], PhD

[1]Office for Healthcare Transformation, Ministry of Health, Singapore, Singapore

[2]Institute of Mental Health, Singapore, Singapore

[3]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[4]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Corresponding Author:
Creighton Heaukulani, PhD
Office for Healthcare Transformation
Ministry of Health
1 Maritime Square
#12-10 Harbourfront Centre
Singapore, 099253
Singapore
Phone: 65 66793209
Email: creighton.heaukulani@moht.com.sg

Abstract

The collection of data from a personal digital device to characterize current health conditions and behaviors that determine how an individual's health will evolve has been called digital phenotyping. In this paper, we describe the development of and early experiences with a comprehensive digital phenotyping platform: Health Outcomes through Positive Engagement and Self-Empowerment (HOPES). HOPES is based on the open-source *Beiwe* platform but adds a wider range of data collection, including the integration of wearable devices and further sensor collection from smartphones. Requirements were partly derived from a concurrent clinical trial for schizophrenia that required the development of significant capabilities in HOPES for security, privacy, ease of use, and scalability, based on a careful combination of public cloud and on-premises operation. We describe new data pipelines to clean, process, present, and analyze data. This includes a set of dashboards customized to the needs of research study operations and clinical care. A test use case for HOPES was described by analyzing the digital behavior of 22 participants during the SARS-CoV-2 pandemic.

(J Med Internet Res 2021;23(3):e23984) doi:10.2196/23984

KEYWORDS

digital phenotyping; eHealth; mHealth; mobile phone; phenotype; data collection; outpatient monitoring; machine learning

Introduction

We are at an age in health care where we have much data at our disposal, including the high penetration of digital electronic medical records and advanced techniques available for their analysis [1]. It is also well accepted that lifestyle characteristics, including activity, stress level, social interactions, and environment, are significant determinants of health outcomes [2,3]. Although estimates vary, it has been argued that *lifestyle choices* exceed the impact of *health care received* as a determinant of premature death [3].

It has been highlighted by Onnela [4] that the wide adoption of smartphones and the increasing use of wearable devices open up a new vista of characterizing both current health conditions and the ongoing behaviors that will determine how an individual's health will evolve. As examples of these new data sources, we can readily measure physical activity, heart rate, heart rate variability, temperature, sleep, *sociability* (amount of human interaction), and smartphone usage (amount and duration of use, type of use, and the way a screen is tapped and scrolled). The approach of using personal digital devices to capture these data sources, and hence characterize an individual *in situ*, has been called *digital phenotyping* [5]. The use of digital phenotyping both complements and extends the use of traditional

home monitoring (eg, blood pressure measurements) in telemedicine by offering continuous measurement during normal activities and everyday living. We have developed a general-purpose digital phenotyping platform called Health Outcomes through Positive Engagement and Self-Empowerment (HOPES), which integrates data from wearable devices and a broad set of smartphone sensors, provides an array of methods to inspect that data, and binds everything together into a platform with a comprehensive privacy and security model. The platform was developed in conjunction with a clinical study for schizophrenia. In what follows, we provide an overview of the HOPES platform and demonstrate its first use.

Digital Phenotyping

The collection of data from a personal digital device can be used to encourage healthy behaviors; an example is the Singapore Health Promotion Board National Steps Challenge [6]. Data collection is sometimes combined with coaching or nudges for general wellness [7] or to monitor and improve an existing diagnosed condition [8,9]. We were originally inspired by the potential use of digital phenotyping to monitor and treat *mental health* conditions such as depression and schizophrenia. Several notable studies, including a study at Northwestern University, have shown the correlation of mobile phone sensors with depressive symptom severity [10]; a recent study at King's College London showed the feasibility and acceptability of the extended use of wearable devices and smartphones in patients with schizophrenia [11] and the use of digital phenotyping for relapse prediction in schizophrenia [12]. The commercial world has also taken notice, and a number of start-ups have formed based on these technologies [13-15].

In addition, digital phenotyping is also being applied to address a diverse range of diseases, such as asthma, maternal health, cancer, and dermatology [9,16-18]. Recent experiments on the use of wearable devices such as the Oura ring [19] and the Fitbit wrist band [20] have been applied to measure participants' parameters during the SARS-CoV-2 pandemic. We illustrate our observations related to the pandemic (using Fitbit) in the *Example Analysis* Section.

Digital phenotyping has the potential to supplement, and in some cases replace, standard clinical processes in data gathering and patient monitoring by virtue of the following attributes:

1. Productivity and cost: passive monitoring can be efficient for both the provider and patient, compared with traditional clinical visits or scheduled telemedicine encounters.
2. Latency: passive monitoring may enable relatively quick responses from health providers, for example, allowing for actions from a case manager within a day versus a week or longer for a visit.
3. Sensitivity: several variables such as resting heart rate or sleep parameters are not easy to measure in the clinic, and device monitoring can be more effective than subjective patient reporting or inconvenient manual processes. The emergence of low-cost consumer devices has been shown to be sufficiently accurate for several purposes [21].
4. More parameters: while in the past we have been limited to infrequent interview questions and scales, we now have the potential to monitor a wider variety of the subjects' parameters, such as location, sleep, motion, heart rate variables, and instantaneous manual responses. These can be measured simultaneously, efficiently, and reliably.

A much-discussed concern is how well such techniques will be accepted and complied with by patients or consumer participants. This involves ease-of-use considerations by both the user and the provider. Another major concern revolves around data security and privacy preservation. These two aspects have been primary motivators in our design choices and investigations.

Clinical Study on Digital Phenotyping in Schizophrenia

The HOPES platform was designed, developed, and refined concurrently to support clinical studies. The HOPE-S (Health Outcomes via Positive Engagement in Schizophrenia) study [22] was launched in November 2019. HOPE-S is an observational study of individuals with schizophrenia who were recently discharged from a psychiatric hospital. The aim of this study is to determine whether digital phenotyping data are associated with clinical and health utilization outcomes. Key events recorded over the 6-month observation period include readmission, outpatient nonattendance (ie, defaults), and unscheduled service use, such as emergency department attendance and mobile crisis team activations. The primary study outcomes are the ability to predict relapse and/or readmission within 6 months, with secondary outcomes being the associations between digital phenotyping data and health care use, psychiatric symptom severity, and functional status assessed during research visits. Ethics approval was granted by Singapore's National Healthcare Group Domain Specific Review Board (reference no.: 2019/00720). To promote the use of digital wearables among our population and to provide incentives to patients to join our study, we offered each participant a Fitbit Charge 3 free of charge.

The first phase of the HOPE-S study is observational. During this phase, we examine the deployment, feasibility, and acceptability of a wide range of digital sensors while performing the analyses required to assess the outcomes described above. In this process, we have been collecting large amounts of data for our analyses. These data will subsequently be used to develop machine learning algorithms to *predict* changes in symptom severity and other important clinical outcomes, as opposed to merely analyzing associations. During the subsequent phase of the study, we will deploy interventions such as early warnings of relapses, which will allow pre-emptive steps to be taken to prevent participant relapse or rehospitalization.

HOPES: A General-Purpose Platform for Digital Phenotyping

HOPES is based on, and extends, the existing *Beiwe* platform [23,24]. Our contributions include the following:

1. The integration of wearable devices, where we have experimented with both wrist and ring devices;
2. The use of a wide range of sensors on the smartphone.
3. An efficient onboarding method for participants.
4. A suite of user interfaces including data collection and quality management tools, clinical summarization

dashboards, and general-purpose research dashboards for use in exploratory data analysis and building anomaly detection algorithms.

5. Assurances for data security and the preservation of user privacy.

The platform is designed to be reliably deployed at scale and makes use of both public clouds and controlled on-premise computing infrastructure. We recognize the broad spectrum of potential applications beyond mental health and the growing set of digital sensors and their capabilities that may be appropriate for different applications. Therefore, we designed HOPES to be flexible and extensible to accommodate new devices and sensor integration, and new data dashboards.

Although the data collected during the HOPE-S study are rich, they are also noisy and incomplete as is expected when dealing with real human behavior and varying data reliability among sensors. To address these challenges, we have developed a *data collection dashboard* and multiple data visualization and exploration tools, which have proven invaluable for monitoring and ensuring participant compliance on a daily basis in the research study. We have also developed a feature engineering pipeline to construct useful insights for the HOPE-S study and to compensate for various shortfalls in the raw data. These dashboards have been found to be easy to use by research coordinators involved in the HOPE-S study, who have been able to easily recognize problems and contact the participant if their data are not being received. We also illustrate the dashboards that our data scientists have used to look for patterns and an *anomaly detection dashboard* that raises alerts on irregularities in the data. All the data are then fed into downstream statistical analyses and our ongoing development of predictive machine learning algorithms. At this stage, our anomaly detection dashboard implements common statistical routines for anomaly detection in time-series data. The development of an effective relapse prediction algorithm is an ongoing subject of this study.

The remainder of this paper is organized as follows: In the section *Existing platforms* we review several existing open source digital phenotyping platforms, highlighting their respective strengths and weaknesses. In the section on *The HOPES Platform and Its First Use in the HOPE-S Study*, we describe the overall architecture of the HOPES platform. In the section on *Dashboards, User Iterfaces, and Data Analysis*, we describe the enhancements to *Beiwe* that the HOPES platform provides, guided by the requirements of the HOPE-S study and other planned future uses (including for purposes beyond mental health). In the section *Example Analysis* we show an early and simple example of the use of our collected data on 22 participants in which we compare user data before and after Singapore's SARS-CoV-2 *lockdown* went into effect. In the section *Conclusions*, we provide some overall conclusions that can be drawn from our experiences with digital phenotyping.

Existing Platforms

There are several existing open source digital phenotyping platforms, including *Beiwe* [23,24], *Purple Robot* [25-27], *AWARE* [28,29], and *RADAR-base* (Remote Assessment of Disease And Relapse) [30-32]. Each contains a core smartphone app that performs passive sensor data collection in the background and a server backend in charge of receiving the data. Note that digital phenotyping is not limited to smartphones; indeed, wearables also provide some significant differentiated capabilities, and there are other sources such as fixed detectors. Some platforms such as *Beiwe* and *RADAR-base* also support active data collection in the form of surveys and some capture data from wearable devices, such as wrist- or arm-wearable devices, by providing a common data interface.

From our assessment, Purple Robot has the most complete coverage of Android sensors and features among the platforms we reviewed. The user can select which sensors to turn on and set the data sampling frequency, however, the platform does not support the iPhone Operating System (iOS). AWARE supports both Android and iOS and has nearly full coverage of Android sensors and features. Similar to Purple Robot, AWARE also allows the user to configure sensors and features. RADAR-base has recently added iOS support and uses both passive (phone use and sensors) and active (survey and questionnaire) data collection. Although it covers fewer phone features and sensors than the Purple Robot and AWARE, it has a very attractive user interface and a very robust system for surveys and questionnaires. *Beiwe* is a smartphone-based digital phenotyping research platform that supports both Android and iOS and has a decent coverage of phone sensors and features. Moreover, the platform supports active feature collection from simple surveys. Apart from the data collection backend that receives data from participants' phones, *Beiwe* also has a backend for data analytics.

We have based the framework for the HOPES platform on *Beiwe* for several reasons. First, *Beiwe* supports both Android and iOS, a requirement for any generic digital phenotyping platform to be widely adopted. Second, our platform analysis and comparison tests conducted in March 2019 showed that at that time, *Beiwe* was most ready to deploy. Our decision was also based on our review of a number of Git repositories and publications as well as previous practical applications of the platforms in clinical studies and trials. We chose the Fitbit wrist device to access data beyond the smartphone sensors after conducting a technical and usability comparison of several popular devices on the commercial market. Specifically, we compared Fitbit Charge 3, Huawei Honor A2, Xiaomi Mi Band 3, Actxa Spur+, and HeyPlus. We found that Fitbit was distinguished by ease-of-use, battery life, and reliability, and it has been validated to be reasonably accurate against gold standard devices for the measurement of sleep [21]. We also evaluated a number of external sleep measurement devices (such as mattress pads) but did not find them suitable for our purposes.

The HOPES Platform and Its First Use in the HOPE-S Study

To support large-scale data aggregation of wearables, mobile phones, and other data sources, we defined a set of requirements and then built our platform to be secure and scalable. Building on top of the existing *Beiwe* platform, we created the HOPES platform by expanding the functional capabilities for easier participant onboarding, enhanced data collection monitoring,

optimized data uploading, extended security features, expanded data processing and analytics pipeline, and a scalable deployment architecture. The goal was to obtain easy and secure onboarding, almost unlimited scaling, high operational security, and improved privacy assurance. Although we were immediately driven by meeting the strict requirements for the HOPE-S study, along the way we became aware of expanded requirements for a wider range of participant monitoring requirements. We took these requirements into account in our architecture and design, so we would be ready for further deployments. In this section, we describe the platform requirements, our resulting HOPES system architecture, the features collected for the HOPE-S study, the enhancements we made to the Android app, the platform backend, and the security protocols. We provide our motivation and a high-level description, leaving further details and information about miscellaneous improvements to Multimedia Appendix 1.

Platform Requirements

The HOPES platform is designed to be a reliable, low-maintenance digital phenotyping collection and aggregation platform. It is designed to support research protocols as well as scale to larger production platforms, including self-service registration. The requirements and their corresponding capabilities are listed in Textbox 1.

To successfully implement such a broad set of requirements, we carefully studied and focused on the user experience for onboarding new participants and built a platform that leverages the best software engineering, design principles, and cloud architecture capabilities.

Textbox 1. Health Outcomes through Positive Engagement and Self-Empowerment (HOPES) platform requirements.

Requirements and implementation capabilities:

Simple user onboarding

- Precreation of user identities and anonymization factors
- Preprinted Quick Response code onboarding sheets
- Ability to migrate participants to new phones (if their current phones are not usable for a study) while maintaining study data integrity and privacy
- Simple onboarding literature and packaging in *gift pack* format
- Wide platform support, Android (with and without Google services), iPhone Operating System
- Ability to be totally passive with zero user interaction after setup
- Preparation for self-service onboarding in the future

User data collection and privacy

- All data deidentified (no personally identifiable information)
- Per-participant encryption keys
- Per-participant random credentials
- Mapping between participant ID and deidentified study ID securely retained but only made available to authorized clinicians
- Secure data backup and archiving

Data security end-to-end

- Data encrypted while in cloud storage environment
- Data decrypted, but still deidentified and obfuscated where appropriate, in data analytics pipeline on-premise

Wearable support that is scalable and secure

- Precreation of wearable device accounts
- Wearable cloud accounts deidentified using study ID
- Wearable data automatically encrypted with user's password
- Server-less functions to periodically collect and archive user data

Infrastructure, scale, and operational security

- Two-factor authentication for all participants, including certificate and credential authentication
- Rotating credentials
- Data collection dashboard
- HOPES work/ticket queue for monitoring alerts/logs/events
- Distributed Denial-of-Service and web application firewall protection
- Elastic scale at all levels
- Isolation of functions across private virtual private local area network
- Separation of administrative and data upload interfaces
- Private virtual private network for administration
- Separation of data upload application programming interface and data management
- Restricted access controls
- Automated repeatable deployment

Data analytics

- Data downloaded on-premise into secure workspace for analytics or clinical use
- Multistage analytics processing pipeline
- Anomaly detection dashboard
- Data exploration dashboard

Security standards

- Secure development process
- Automated patching
- Additional requirements from Singapore security and Information Technology standards

Expanded data collection support for social media metadata

- Support for deidentified metadata for WhatsApp text and audio/video messages on Android
- Phone text messages

Study clinical support for easy clinical management

- Daily deidentified data collection dashboard emailed to study researchers and clinicians to monitor study compliance
- Encrypted deidentified clinician dashboard accessible to clinicians

Overall System Architecture

The high-level solution architecture of HOPES, as used in the HOPE-S study, is shown in Figure 1. On each participant's smartphone, we installed two apps: the Fitbit app and the HOPES app. Every participant was required to wear a Fitbit watch for a certain portion of the day and night (enough to collect the required data but also allowing removal for charging, showering, etc). Fitbit raw data are collected by the Fitbit app and sent to a Fitbit server (the *Fitbit Cloud*) for processing and computation of high-level features (eg, the estimation of sleep stages). Phone data are collected on the smartphone by the HOPES app and sent to *a data upload server* hosted in a public cloud using Amazon Web Services (AWS). The *data processing backend server*, located at either the Research and Development (R&D) or clinical premises, periodically pulls data from both the Fitbit cloud and the AWS *data upload server* for subsequent processing, as described in the following sections. The data are always deidentified when in a publicly accessible cloud environment, and all transmissions and storage are encrypted. Certain variables, such as location, are also obfuscated at the time of collection for privacy preservation. More details on the solution architecture are provided in the Multimedia Appendix 1.

For backend R&D analytics, we developed a set of data processing pipelines and various dashboards for monitoring, visualizing, and analyzing data (Figure 2). The data processing pipelines clean (manage missing, duplicated, and erroneous data), convert, and reorganize data into more usable forms. These dashboards are used by research coordinators and clinicians, researchers, data analysts, and technical team members involved in the conduct of the study. A general-purpose research dashboard supports exploratory analytics. In each case, roles and responsibilities determine the access controls for various attributes of the data. Physical controls, supervision, and accountability measures were also deployed to ensure that there was no unauthorized access to data. Further description is given in subsequent sections and more details are provided in the Multimedia Appendix 1.

Figure 1. Overall system architecture and data flow diagram for the Health Outcomes via Positive Engagement in Schizophrenia (HOPE-S) study.

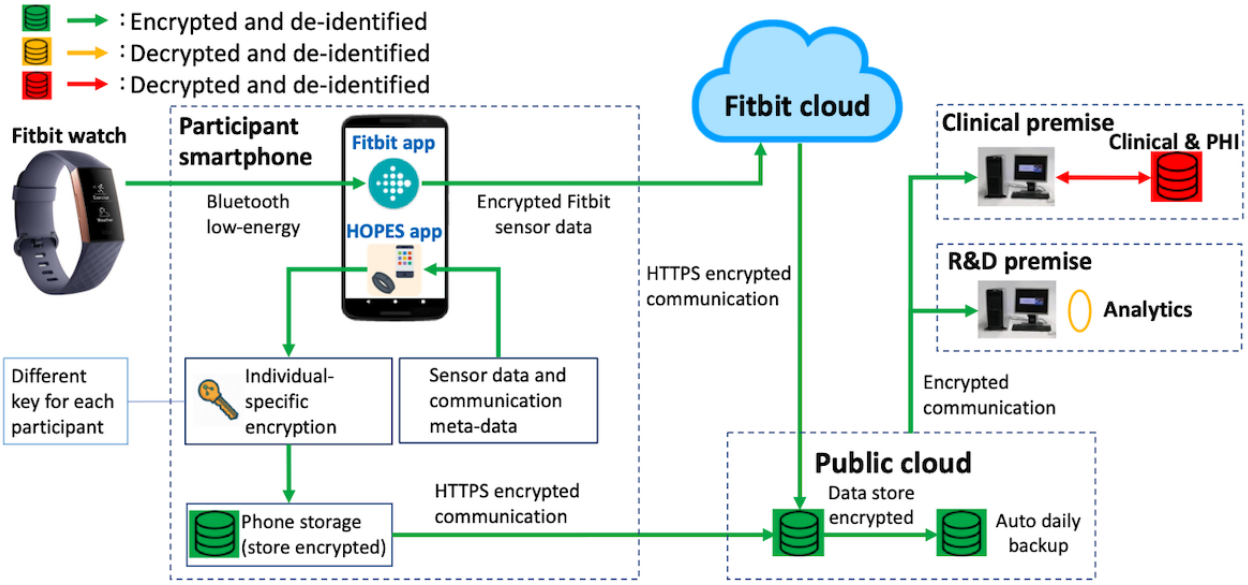
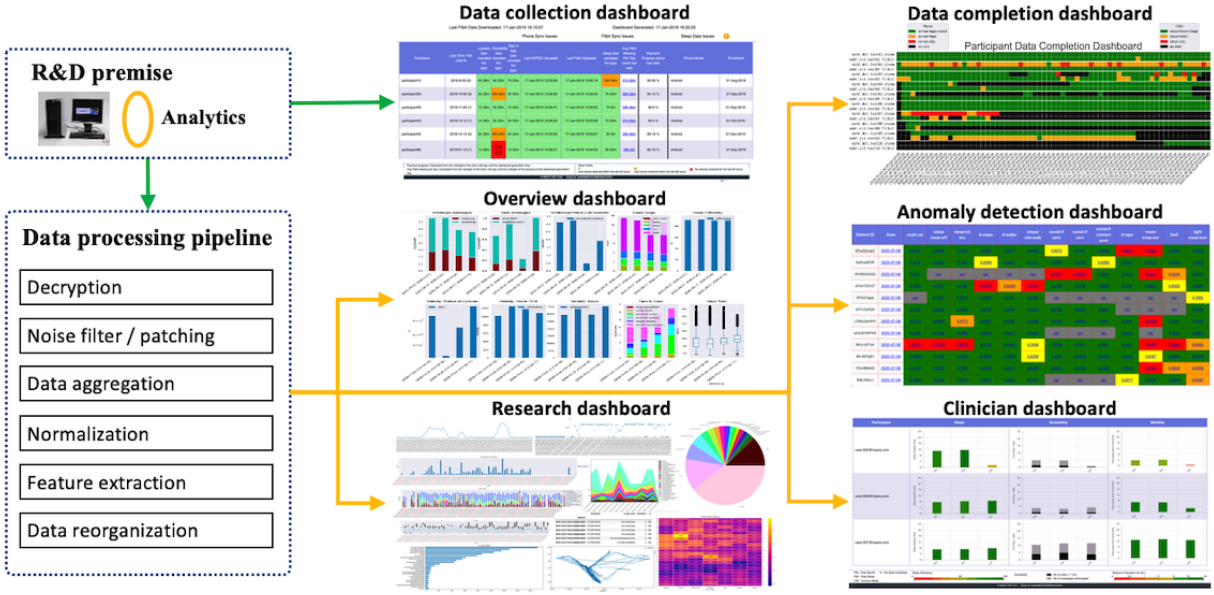


Figure 2. Backend data processing pipelines and dashboards.



Features Supported by the Platform

The following 6 categories of features are obtained from the HOPES smartphone app. In each case, we will indicate *new* if it is a new feature added by us or an enhancement, otherwise, it is an existing feature in the *Beiwe* distribution.

Location

GPS coordinates are used to detect deviations from typical travel patterns and to compute a measure of variance or entropy in the locations visited by a participant. To protect user privacy, the raw GPS coordinates are obfuscated via a random displacement (from the origin), which is unique for every participant.

Sociability Indices (Some Are New)

In our study, changes in a participant's s*ociability*, that is, their communication with others, is estimated from available data. Sociability may be reflected in their activities in various forms of messaging and voice/video communications. The original *Beiwe* app can capture incoming and outgoing phone calls and SMS messages. However, in many countries, most people use free social messaging apps as their primary method for text and voice communication; for example, in Singapore WhatsApp use is dominant. We therefore made use of the Android Accessibility Service Application Programming Interface (API) to acquire message metadata from social messaging apps. So far, we have only implemented this for WhatsApp, but it can be easily extended to other social messaging apps. The duration and timing of mobile service phone calls and WhatsApp calls made and received were recorded. Similarly, the length and timing of SMS and WhatsApp messages sent and received were also recorded. Importantly, for privacy protection, we never record or transmit any content of any communication, and we hash the identity or contact number of the counterparty.

Finger Taps (New)

Taps provide two types of information that may be related to a person's health. The speed at which a person taps may give a hint of their neuropsychological function [33]; for example, a fatigued person may tap more slowly or some diseases may cause small, uncontrollable movements. There is also some evidence that finger taps may be used to detect depression [34]. The apps a person uses (determined from their taps) may also give an indication of their status and behavior. For example, a patient with a mental illness who is relapsing might be found to have significantly altered communications, reflected in the number and speed of taps made in the various apps. We sought to capture typing error rates that could be affected by physical or mental conditions. We can determine this from how often the delete or backspace key on the keyboard is tapped. To measure tapping speed, we also need to know whether the person is typing on the keyboard or navigating in a social messaging app. The characteristics and metadata of finger taps on the phone screen were recorded, such as the number and timestamps of taps into apps, different key strokes (from the enter key, delete key, backspace key, alphabet keys, number keys, and punctuation keys), and the group categorization of the tapped apps are also recorded. As a privacy-preservation measure, captured keystrokes are converted into a type token (such as *alphabetic*, *numerical*, and *punctuation*). The app only stores and downloads the type token, and the specific keys that are struck are not recorded.

Motion Information (Some Are New)

Accelerometer, gyroscope, magnetometer (new), and pedometer (new) data are recorded to check whether the phone is being moved or is motionless. This information can help determine the amount of phone use and can be correlated with other data collected by wearable devices (sleep, activity, etc).

Phone States

The app can record the Wi-Fi state, the Bluetooth state, and the power state (screen on/off and power-down event) of the phone. The Wi-Fi and Bluetooth scan results can, to some extent, provide information about the location of the device, especially when the GPS location is not available. However, these data

are sensitive and need to be deidentified and encrypted. The power state feature is usually combined with other features, such as taps, to determine the usage behavior of the phone by the participant.

Ambient Light (New)

The app can record the intensity of ambient light through the smartphone's built-in light sensor (not the camera). This could detect, for example, whether a participant goes out or sleeps in a comfortable sleeping environment; studies have suggested a correlation between a patient's mental health and their preferred environmental lighting [35]. As sleep and heart rate are important indicators of mental health status, we recorded the following 3 categories of features from the Fitbit wearable (obtained directly from the Fitbit cloud).

Sleep

Sleep information during the day and night was recorded, including a breakdown of different sleep stages with time stamps.

Steps

The total number of steps in time intervals specified by Fitbit.

Heart Rate

The number of heartbeats in time intervals specified by Fitbit. Approximations of other measures of interest, such as heart rate variability, can be computed from heart rate data.

For the HOPE-S study, we captured the following features: location, sociability indices, finger taps, accelerometer, power state, ambient light, sleep, steps, and heart rate.

Backend Data Processing Pipeline

We have rebuilt the *BBS* backend in Python 3 to systematically process data files, reformat the raw data, and extract high-level features. A considerable amount of feature engineering is being performed on the backend to clean the data, correct data shortcomings, combine different data sources into joint features, and feed various downstream machine learning systems. For example, upon consultation with our clinical partners, we constructed high-level features that are likely to provide useful signals regarding the mental health of the participants in the HOPE-S study. Our current analyses in the study make use of time series of daily or hourly samples of intuitively identified measurements from sleep, steps, heart rate, location, and sociability indices. Some examples include daily totals of the number of hours of sleep, steps, and communications initiated and received. Constructing such features is often necessary in situations with small amounts of or noisy data. For example, when no sleep data are recorded by the Fitbit for a whole day, it is not clear whether the participant did not sleep or whether they just did not wear the Fitbit to bed. We can resolve this ambiguity by looking at the heart rate measurements, which are recorded continually while the Fitbit is worn. If heart rate data are missing for more than an allotted allowance, we can reasonably assume that the participant was not wearing the Fitbit during sleep. As another example, we have developed an Android app grouper that uses information from the Google Play Store to classify all apps into 7 classes defined by us: social messenger, social media, entertainment, map navigation, utility tools, games, and Android systems (other vendor-specific or system apps that cannot be found in the Play Store). This class information is used in the taps data features when classifying a user's phone activity (eg, *in social media apps*, *in gaming apps*, etc). In summary, this step bridges the gap between data collection and common downstream machine learning modules. Details on the data processing pipeline, high-level feature extraction, and the seven classes of the app grouper are provided in the Multimedia Appendix 1.

We note that although we endeavor to correct *ambiguities* in the data collected by the platform (such as in the example above clarifying truly sleepless nights), we do not make efforts to *impute* missing data. Imputation is required for certain analyses, such as those involving GPS measurements [36]. However, it should be noted that the best imputation method depends on the goals of a particular study.

We are also aware that features that are provided by device manufacturers, such as the pedometer, heart rate, and sleep, are derived using proprietary algorithms that are likely to change over time and are not standardized, nor typically scientifically validated. These features may contain biases or inaccuracies that can affect subsequently trained statistical models. Therefore, our existing data processing pipeline is designed to be flexible enough for researchers to insert on-demand additional steps for data normalization and regularization.

Platform Improvements

We have made many improvements to the Android app and are in the process of extending these improvements to the iOS app. In this section, we will only describe the most significant improvements; other improvements are provided in the Multimedia Appendix 1. We also used two system variations: the *prototype* or development system and the *deployed* system. Some features may be applied to only one of the systems.

Scanning Quick Response Codes for Simple User Registration

To facilitate the user registration process and to allow one-way encryption for better data security, study participant kits were prepared and a single-page onboarding document was generated with all the information necessary to onboard a participant. The process was designed for a nontechnical self-service onboarding process. Multiple Quick Response (QR) codes were scanned in the deployed system. They include information on certificate-based authentication to further strengthen security via host verification. The *Additional App Enhancement* Section of the Multimedia Appendix 1 provides details on QR registration.

Data Compression

To scale the system up to a very large number of users, we need to reduce the utilized communication bandwidth as much as possible. We have therefore added an option when creating a *study* to compress the data before sending it to the server, which may be selected on the backend console by checking the *enable compression* checkbox. Note that data compression is applied

before data encryption. This feature was only implemented in *the prototype system*.

Security Enhancements

The HOPES platform is redesigned on top of *Beiwe* to ensure data confidentiality, data integrity, and high availability, and to enable system auditing and user authentication. The design also supports large-scale deployments with a distributed pipeline. Finally, the design emphasizes a separation of these duties throughout the architecture to minimize the risk of data breaches and to preserve data privacy throughout the lifecycle of a study. In the original *Beiwe* platform, data are decrypted in the data collection server and re-encrypted using the study key. This poses a certain amount of risk because the data collection server directly faces the public internet. In our HOPES platform, data are encrypted at all times while on the phone and in the data collection infrastructure and are only decrypted in clinical or R&D premises. The decryption key is only accessible from clinical or R&D premises; therefore, in principle, the data are not decryptable on the phone or in the data collection infrastructure. Data are only reidentified when needed for qualified clinical purposes and only by clinical staff.

Dashboards, User Interfaces, and Data Analysis

Ensuring complete data collection is important. A variety of issues can result in not receiving data as expected, including technical failures, participants not adhering to the guidelines on device usage, or participants failing to wear their device. Monitoring this process is particularly challenging at scale. Therefore, we created a *data collection dashboard* (Figure 3) to facilitate the monitoring of the data collected.

The data collection dashboard is populated using the metadata extracted during the downloading phase of the Fitbit and phone data. The AWS Lambda function (which is set to trigger every 5 minutes) is set up to retrieve these data from their respective S3 buckets and create an HTML file. To fill the dashboard to ensure that the participants comply with the study requirements, the following data types are observed and closely monitored: location, sociability, taps in app, last HOPES upload, last Fitbit upload, and sleep. Color codes denote the data collection status: red meaning need to take an action, orange meaning need to closely monitor, and green meaning normal.

The data collection dashboard does not require decrypted data and is thus constructed before decryption. As a result, it can be hosted on an upload server with little security risk. However, it does not show the full historical data completion status, which is sometimes needed. Hence, we developed the *data completion dashboard*, which is described in detail in the Multimedia Appendix 1.

Figure 3. The data collection dashboard shows the data uploading status of all participants.

Data collection dashboard - HOPES

Last Fitbit Data Downloaded: 04-Aug-2020 08:53:32

Dashboard Generated: 04-Aug-2020 10:39:34

Phone Sync Issues

Fitbit Sync Issues 1 1
user.002@hopes.com
user.004@hopes.com

Sleep Data Issues 1 1
user.002@hopes.com
user.004@hopes.com

Participant	Last Clinic Visit (visit #)	Location (last recorded hrs ago)	Sociability (last recorded hrs ago)	Taps in App (last recorded hrs ago)	Last HOPES Uploaded	Last Fitbit Uploaded	Sleep (last uploaded hrs ago)	Avg Fitbit Wearing Per Day (since last visit)	Payment Progress (since last visit)	Phone Model	Enrollment
user.006@hopes.com	2020-06-18 (2)	8h 39m	9h 39m	14h 39m	04-Aug-2020 10:16:32	03-Aug-2020 22:45:00	24h 51m	22h 37m	96.05 %	Android	24-Mar-2020 10:50:03
user.003@hopes.com	2020-07-17 (3)	1h 39m	1h 39m	1h 39m	04-Aug-2020 09:57:29	04-Aug-2020 00:22:04	26h 59m	22h 36m	89.49 %	Android	29-Jul-2020 17:33:00
user.009@hopes.com	2020-06-26 (2)	21h 39m	21h 39m	21h 39m	03-Aug-2020 22:35:40	04-Aug-2020 06:42:44	27h 32m	21h 19m	90.67 %	Android	30-Mar-2020 11:46:25
user.005@hopes.com	2020-07-23 (3)	4h 39m	33h 39m	4h 39m	04-Aug-2020 09:49:55	04-Aug-2020 07:20:55	19h 50m	22h 0m	91.67 %	Android	16-Mar-2020 11:56:06
user.007@hopes.com	2020-07-23 (3)	3h 39m	89h 39m	22h 39m	04-Aug-2020 10:13:25	04-Aug-2020 08:17:36	23h 57m	22h 0m	91.67 %	Android	25-Mar-2020 10:09:44
user.008@hopes.com	2020-07-30 (3)	2h 39m	17h 39m	3h 39m	04-Aug-2020 10:09:22	04-Aug-2020 08:41:28	4h 13m	24h 0m	96.36 %	Android	26-Mar-2020 13:38:50
user.001@hopes.com	2020-07-09 (3)	4h 39m	4h 39m	4h 39m	03-Aug-2020 14:40:43	04-Aug-2020 08:42:46	7h 5m	22h 26m	94.23 %	Android	03-Aug-2020 14:40:31
user.002@hopes.com	2020-07-13 (3)	41h 39m	18h 39m	107h 39m	03-Aug-2020 21:20:23	30-Jul-2020 12:41:59	144h 49m	16h 54m	72.51 %	Android	09-Mar-2020 11:02:48
user.004@hopes.com	2020-07-14 (3)	1h 39m	85h 39m	1h 39m	04-Aug-2020 09:53:45	31-Jul-2020 20:17:35	87h 23m	20h 6m	84.77 %	Android	11-Mar-2020 13:58:53

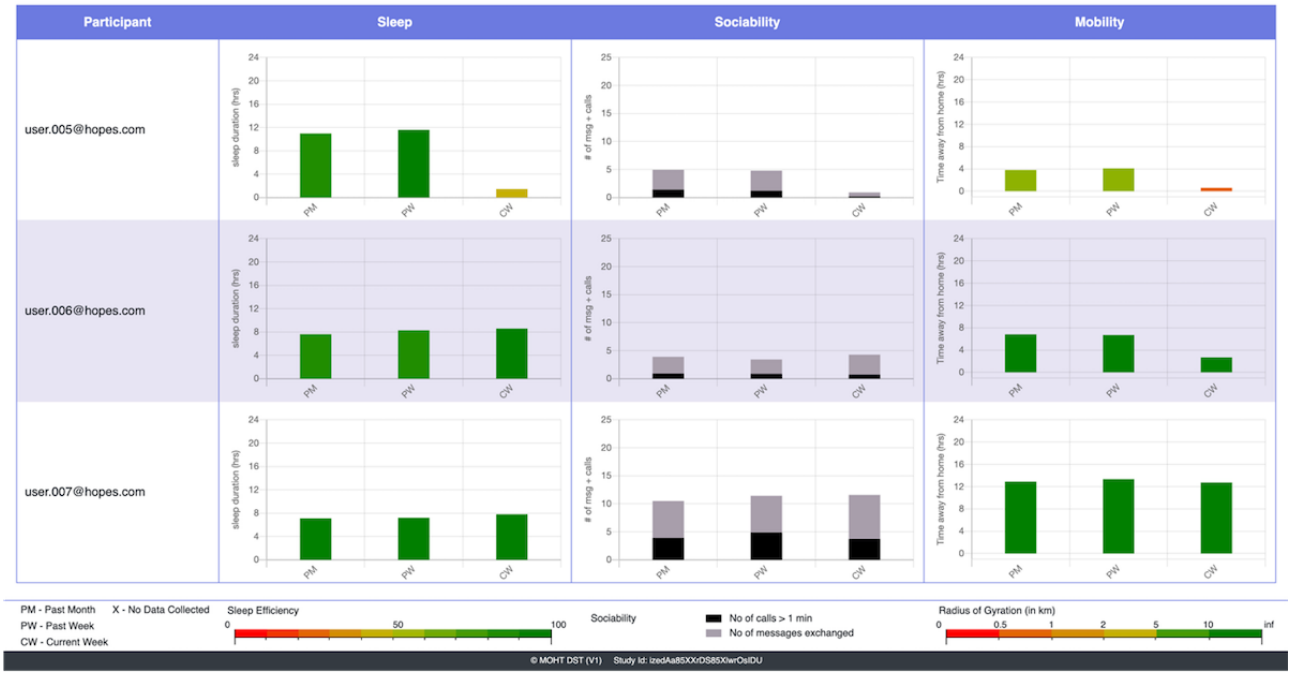
Data Visualization Toolkit

We developed a *data visualization toolkit* to visualize and explore the collected data. The toolkit can also perform some basic statistical analyses, such as the comparison of features between defined date ranges. For further details on the usage and capability of the data visualizer, see the Multimedia Appendix 1.

Clinician Dashboard

The *clinician dashboard*, illustrated in Figure 4, is designed for clinicians to preview general trends in participants' digital phenotyping data and may be useful during clinical encounters. On the basis of previous studies and the observations of our clinical partners, we decided to report sleep, sociability, and mobility data for the current version of the clinician dashboard.

Figure 4. The clinician dashboard shows a preview of general trends in patient biomarker data.



Sleep is plotted based on total sleep duration and sleep efficiency; the latter is depicted by color. Sociability is plotted using the number of messages exchanged and the number of calls for a duration of more than 1 minute. Mobility is based on the time away from home (time spent away from sleeping location) and the radius of gyration (maximum distance traveled from home). These graphs are drawn based on averages over 3 timeframes: the current week is 7 days before 0:00 AM of the current day, the past week is 7 days before the current week, and the past month is 30 days before the past week. An example further explaining the clinician dashboard can be found in the Multimedia Appendix 1.

Anomaly Detection Dashboard

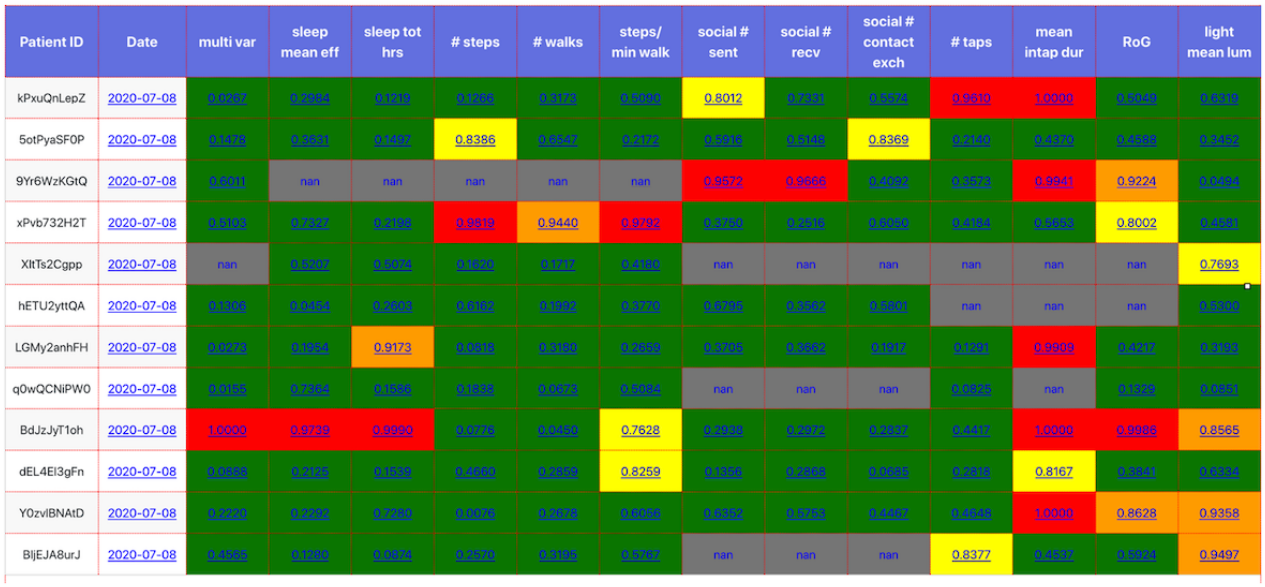
To support a wide variety of applications attempting to analyze and identify interesting changes among the many features being collected by the platform, we implemented a generic purpose *anomaly detection system and dashboard*. The system comprises several anomaly detection algorithms on the backend that report their findings via a dashboard. The dashboard is designed to create alerts about possible irregularities arising in the digital phenotyping data each day.

There are many machine learning approaches to anomaly detection in time-series data. One approach is to train a time-series model on historical data and compare new data with forecasts from this model, *scoring* the predictions based on how *good* or *bad* they are. For example, a simple scoring mechanism compares the empirical distribution of the *residuals* (ie, the errors of the fitted model's predictions on the training set) to the realized prediction error on new data.

We have experimented with several time-series models, including the broad class of autoregressive integrated moving average models [37] and the class of *Gaussian processes* [38], fitting them to a subset of digital phenotyping features that were initially selected as important for our HOPE-S study (see the Multimedia Appendix 1 for details of the features). We note that these two choices of models are able to capture *periodic effects*, which are important for our HOPE-S study, as participants' behaviors may change markedly on the weekends. Selecting the most appropriate model depends on the data and the application at hand. We train the models every day on all past data and compute the predictions of the digital phenotyping features for the next day. At the end of the following day, the realized digital phenotyping features were compared with the predictions and scored, and these scores were transformed to be interpreted as *the probability that the observed data is an anomaly*. Therefore, the final score is a number between zero and one, where higher values constitute alerts.

In Figure 5, we display an example of what the anomaly detection dashboard looks like on a given day. Each row corresponds to a participant, and each column corresponds to a different anomaly detection score. The participant's identifier and the last date their scores were successfully updated are displayed, along with the anomaly scores for each feature. The score from a multivariate model is also displayed, which may capture interdependencies between features that affect whether a measurement is anomalous. For example, major disruptions in sleep naturally coincide with periods of long-distance travel (a large radius of gyration). Note that the cells are colored according to the severity of the scores.

Figure 5. The anomaly detection dashboard with a visualization of the scores from a collection of anomaly detection models.



Patient ID	Date	multi var	sleep mean eff	sleep tot hrs	# steps	# walks	steps/ min walk	social # sent	social # recv	social # contact exch	# taps	mean intap dur	RoG	light mean lum
kPxuQnLepZ	2020-07-08	0.0267	0.2984	0.1219	0.1298	0.3129	0.5090	0.8012	0.7331	0.5574	0.9810	1.0000	0.5049	0.6319
5otPyaSF0P	2020-07-08	0.1478	0.3631	0.1457	0.8386	0.6547	0.2172	0.5916	0.5149	0.8369	0.2140	0.4279	0.4389	0.3452
9Yr6WzKGtQ	2020-07-08	0.6011	nan	nan	nan	nan	nan	0.9572	0.9666	0.4089	0.3573	0.9841	0.9224	0.0694
xPvb732H2T	2020-07-08	0.5103	0.7327	0.2196	0.9819	0.9440	0.9792	0.3750	0.2515	0.6050	0.4164	0.9553	0.8002	0.4581
XltTs2Cgpp	2020-07-08	nan	0.5207	0.5074	0.1620	0.1717	0.4180	nan	nan	nan	nan	nan	nan	0.7693
hETU2yttQA	2020-07-08	0.1306	0.0464	0.2603	0.6182	0.1982	0.3770	0.6796	0.3562	0.5901	nan	nan	nan	0.5300
LGMy2anhFH	2020-07-08	0.0273	0.1954	0.9173	0.0818	0.3160	0.2669	0.3706	0.3962	0.1917	0.1291	0.9909	0.4217	0.3793
q0wQCNiPW0	2020-07-08	0.0155	0.7364	0.1586	0.1838	0.0673	0.5084	nan	nan	nan	0.0825	nan	0.1329	0.0851
BdJzJyT1oh	2020-07-08	1.0000	0.9739	0.9990	0.0776	0.0499	0.7628	0.2938	0.2372	0.2937	0.4417	1.0000	0.9986	0.8565
dEL4El3gFn	2020-07-08	0.0688	0.2125	0.1539	0.4603	0.2859	0.8259	0.1356	0.2868	0.0085	0.2018	0.8167	0.3841	0.6334
Y0zvlBNAtD	2020-07-08	0.2320	0.2293	0.2280	0.0076	0.2078	0.6056	0.6352	0.5753	0.4867	0.4648	1.0000	0.8628	0.9358
BIjEJA8urJ	2020-07-08	0.4569	0.1282	0.0874	0.2570	0.3195	0.5767	nan	nan	nan	0.8377	0.4537	0.5024	0.9497

Although this dashboard is mainly used for research at this point, if reliable anomalies are detected, they can be promoted to the clinician dashboard. In the context of our HOPE-S study, it was shown that digital phenotyping signals from patients with schizophrenia exhibit a measurable increase in anomalies in the period leading up to a relapse event [11]. However, as an unsupervised learning problem, the performance of an anomaly detection routine is dependent on the context of its application, and users will likely have to adjust the underlying algorithm to suit their needs. Therefore, we made the dashboard modular, where the anomaly detection routine can be replaced on the backend with customized routines without affecting its exposure to the user-facing dashboard.

Example Analyses

Data Completion Rate Analysis

The completeness of the data collected in the HOPE-S study depends on the technical stability of the platform and participant compliance. There will always be situations that are difficult to anticipate and can cause losses of data (eg, a few participants had their Fitbit wrist bands broken in the middle of the study, and while waiting for the replacement band to arrive, they were unable to wear the device).

Table 1 shows the data completion rates for each low-level feature. The rate is computed as the number of days with feature data divided by the number of days enrolled in the study. For phone features such as call log and SMS log, which can have no data if the participant really has no call/SMS during that day, we checked the presence of empty, time stamped feature files stored by the platform to determine whether that feature is being successfully collected. Our researchers and clinical partners generally felt that the overall completion rate was satisfactory.

Figure 6 provides a dynamic, graphical overview of the data completeness for each participant. Each participant had 2 rows in the display: the first for their phone features and the second for their wrist device features. Each square represents the completion status for a single day (refer to the legend for information on color coding). This dashboard has proven valuable to our researchers and clinical partners when following up with study participants to improve compliance and to quickly resolve any technical issues that may arise.

Table 1. Data completion rate on 22 participants that have completed the study.

Raw feature name	Data completion rate (%)
Accel	87.0
Call log	94.6
Power state	94.3
Sociability call log	94.6
Accessibility log	87.2
GPS	93.0
Sleep	87.1
Sociability msg log	94.6
Ambient light	91.3
Heart	93.5
SMS log	94.6
Steps	96.9
Taps log	89.7
Overall	92.2

Figure 6. Data completion overview for the first 22 participants who have already completed the study (odd rows refer to phone data; even rows refer to Fitbit data).

Example of Use: Measuring the Effect of Singapore's Circuit Breaker

In response to the SARS-CoV-2 (COVID-19) pandemic, Singapore imposed a stay-at-home order or *cordon sanitaire*, which is formally called *the 2020 Singapore Circuit Breaker measures* or CB. This lockdown was in effect from April 7, 2020 to June 1, 2020, after which gradual stages of reopening have occurred. During this period, people were required to stay at home as much as possible, avoid nonessential travel and social visits, and maintain social distancing in public. We expect the lockdown to have an effect on some digital phenotyping features. As a test for our digital phenotyping system, we performed and reported a data comparison using 22 participants' data before and after the start of this *CB*.

Table 2 shows a subset of features that show statistically significant differences before and after the CB began on April 7, 2020. Not surprisingly, as people were required to stay at home, the time at home has increased, and the number of significant locations visited has decreased. Features related to physical activity (heart rate, steps, and acceleration) also decreased, as might be expected. Both sleep and sleep efficiency decreased among these participants. It is also noteworthy that participants appear to use a fewer number of apps, perhaps because there is no need for some apps such as maps for navigation or those checking bus arrival times; however, it appears that they spend more time in entertainment apps. Moreover, the ambient light indoors is generally dimmer than it is outdoors, therefore, the observed decrease in maximum ambient light is also as expected.

Table 2. Comparison of 6 weeks of digital phenotyping data before (from 45 days before to 3 days before) and after (from 3 days after to 45 days after) the start of Singapore's Circuit Breaker on April 7, 2020.

Feature name	6-week before CB[a] starts, mean (SD)	6-week after CB starts, mean (SD)	*P* value (paired *t* test)	*P* value (Wilcoxon signed rank test)
Smartphone features				
accel_L_std (L: length of the accel. vector)	0.526 (0.325)	0.370 (0.345)	<.001	<.001
accel_ddt_max (ddt: time derivative)	0.00807 (0.0044)	0.00551 (0.0041)	.005	<.001
ambientLight.hourly_max_log1p_lux	2.538 (0.721)	2.190 (0.587)	.001	.001
callLog_Incoming Call	0.533 (0.835)	0.288 (0.560)	.001	<.001
gps-mobility_Hometime / mins	1111 (292)	1328 (143)	<.001	<.001
gps-mobility_SigLocsVisited	1.354 (0.429)	1.220 (0.288)	.01	.02
powerState.hourly_n_screen_on	6.716 (2.77)	5.180 (2.44)	.008	<.01
tapsLog.daily_n_unique_apps	16.10 (4.54)	14.24 (4.47)	.04	.04
tapsLog.daily_n_taps_in_entertainment	293.9 (259)	378.3 (365)	.07	.05
Wrist-wearable features				
steps.daily_n_steps	3921 (2435)	2400 (1807)	.001	<.001
steps.daily_n_mins_walk	113.8 (53.23)	79.78 (54.24)	<.001	<.001
heart.daily_HR_mean	82.89 (10.48)	80.35 (9.04)	.10	.03
heart.daily_HR_min	56.08 (6.85)	55.74 (7.23)	.77	.26
sleep_total_hrs	9.196 (1.94)	8.727 (1.98)	.31	.096
sleep_mean_efficiency	93.27 (3.08)	92.39 (3.38)	.04	.009

[a]CB: circuit breaker.

We compared our results with another study based on Fitbit use: the Health Insights Singapore (hiSG) study [39]. The daily step count decreased by approximately 35% in our study and approximately 42% in hiSG; the minimum heart rate decreased by 1.1 bpm in our study, and the resting heart rate decreased by 1.6 bpm in the hiSG study; sleep efficiency decreased by 0.8% in our study and by 0.2% in the hiSG study. All comparisons between both studies were consistent in demonstrating changes before and after the onset of the CB measures in Singapore.

Conclusions and Future Work

Digital phenotyping is a promising area in health care, but great care and effort is required in designing a system that is easy to use, is safe in terms of data security and privacy, and collects data with sufficient detail and reliability to be useful in research and patient care. We found the *Beiwe* platform to be a suitable base that can be used and extended to create the HOPES platform. Our main extensions have been adding many more sources for data collection, integrating the use of a wearable device, and the development of a large set of participant monitoring and management platforms.

We were also driven by the requirements of a clinical research study for schizophrenia (HOPE-S). This required us to develop significant enhancements in security, privacy, ease of use, and scalability, choosing a careful combination of public cloud and on-premises operation.

We needed to create new mechanisms to clean, process, present, explore, and analyze the massive and diverse data collected when digital phenotyping. These need to serve the needs of clinical research study operations, clinical care, platform developers, and researchers, and hence a range of data pipelines and tools for data analysis have been developed.

Our initial platform is in use in the HOPE-S clinical trial, and interim results will soon be reported. An initial analysis using SARS-CoV-2 as a test case yielded meaningful and expected results consistent with expected lockdown behaviors and was consistent with an independently conducted study in the same country.

Currently, we are considering making the HOPES platform open source for the research community to access. We continue to add features and make adjustments for newer versions of the Android operating system. Simultaneously, we are working on an iOS version of the app to access a wider range of users with Apple devices.

Acknowledgments

The authors thank their colleagues at Singapore's Institute of Mental Health, who have been involved in defining and testing the HOPES system and leading the HOPE-S clinical study. They also thank Fitbit for providing technical assistance on using their cloud API.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Further technical documentation on the solution architecture, App enhancements, added digital phenotyping sensors and features, the system's Fitbit component, the back-end data processing pipeline, and the dashboards.
[DOCX File , 7556 KB - jmir_v23i3e23984_app1.docx]

References

1. Marsch LA. Digital health data-driven approaches to understand human behavior. Neuropsychopharmacology 2020 Jul 12;1(6) Published online ahead of print. [doi: 10.1038/s41386-020-0761-5]
2. McGinnis JM, Williams-Russo P, Knickman JR. The Case For More Active Policy Attention To Health Promotion. Health Affairs 2002 Mar;21(2):78-93. [doi: 10.1377/hlthaff.21.2.78]
3. Schroeder SA. Shattuck Lecture. We can do better--improving the health of the American people. N Engl J Med 2007 Sep 20;357(12):1221-1228. [doi: 10.1056/NEJMsa073350]
4. Onnela J. Opportunities and challenges in the collection and analysis of digital phenotyping data. Neuropsychopharmacology 2020 Jul 17;46:45-54. [doi: 10.1038/s41386-020-0771-3]
5. Torous J, Kiang MV, Lorme J, Onnela J. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. JMIR Ment Health 2016 May 05;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
6. National Steps Challenge. URL: https://www.healthhub.sg/programmes/37/nsc [accessed 2020-08-19]
7. Waddell KJ, Shah PD, Adusumalli S, Patel MS. Using Behavioral Economics and Technology to Improve Outcomes in Cardio-Oncology. JACC: CardioOncology 2020;2(1):84-96. [doi: 10.1016/j.jaccao.2020.02.006]
8. Cote DJ, Barnett I, Onnela J, Smith TR. Digital Phenotyping in Patients with Spine Disease: A Novel Approach to Quantifying Mobility and Quality of Life. World Neurosurgery 2019;126:e241-e249. [doi: 10.1016/j.wneu.2019.01.297]
9. Wright AA, Raman N, Staples P, Schonholz S, Cronin A, Carlson K, et al. The HOPE Pilot Study: Harnessing Patient-Reported Outcomes and Biometric Data to Enhance Cancer Care. JCO Clin Cancer Inform 2018;2:1-12. [doi: 10.1200/CCI.17.00149]
10. Saeb S, Zhang M, Kwasny MM, Karr CJ, Kording K, Mohr DC. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. Journal of Medical Internet Research 2015 Jul 15;17(7):e175. [doi: 10.2196/jmir.4273]
11. Meyer N, Kerz M, Folarin A, Joyce DW, Jackson R, Karr C, et al. Capturing Rest-Activity Profiles in Schizophrenia Using Wearable and Mobile Technologies: Development, Implementation, Feasibility, and Acceptability of a Remote Monitoring Platform. JMIR Mhealth Uhealth 2018 Oct 30;6(10):e188. [doi: 10.2196/mhealth.8292]
12. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela J. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology 2018 Jul;43(8):1660-1666. [doi: 10.1038/s41386-018-0030-z]
13. Mindstrong Health. URL: https://mindstronghealth.com/ [accessed 2020-08-19]
14. Lief Therapeutics. URL: https://getlief.com/ [accessed 2020-08-19]
15. QuantActions. URL: https://quantactions.com/ [accessed 2020-08-19]
16. Jaimini U, Thirunarayan K, Kalra M, Venkataraman R, Kadariya D, Sheth A. "How Is My Child's Asthma?" Digital Phenotype and Actionable Insights for Pediatric Asthma. JMIR Pediatr Parent 2018;1(2):e11988. [doi: 10.2196/11988]
17. Faherty LJ, Hantsoo L, Appleby D, Sammel MD, Bennett IM, Wiebe DJ. Movement patterns in women at risk for perinatal depression: use of a mood-monitoring mobile application in pregnancy. J Am Med Inform Assoc 2017;24(4):746-753. [doi: 10.1093/jamia/ocx005]
18. Spinazze P, Bottle A, Car J. Digital Health Sensing for Personalized Dermatology. Sensors 2019 Aug 05;19(15):3426. [doi: 10.3390/s19153426]
19. Oura Ring. URL: https://ouraring.com/ucsf-tempredict-study [accessed 2020-08-19]
20. Hemphill NM, Kuan MT, Harris KC. Reduced Physical Activity During COVID-19 Pandemic in Children With Congenital Heart Disease. Canadian Journal of Cardiology 2020 Jul;36(7):1130-1134. [doi: 10.1016/j.cjca.2020.04.038]
21. Lee XK, Chee NI, Ong JL, Teo TB, van Rijn E, Lo JC, et al. Validation of a Consumer Sleep Wearable Device With Actigraphy and Polysomnography in Adolescents Across Sleep Opportunity Manipulations. Journal of Clinical Sleep Medicine 2019 Sep 15;15(09):1337-1346. [doi: 10.5664/jcsm.7932]
22. Clinicaltrials.gov. URL: https://clinicaltrials.gov/ct2/show/NCT04230590 [accessed 2020-08-19]
23. Beiwe Research Platform. URL: https://www.beiwe.org/ [accessed 2020-08-19]
24. Onnela-lab. URL: https://github.com/onnela-lab [accessed 2020-08-19]
25. Schueller SM, Begale M, Penedo FJ, Mohr DC. Purple: a modular system for developing and deploying behavioral intervention technologies. J Med Internet Res 2014 Jul 30;16(7):e181 [FREE Full text] [doi: 10.2196/jmir.3376] [Medline: 25079298]

26. Purple Robot - CBITs TECH Web Site. URL: https://tech.cbits.northwestern.edu/purple-robot/ [accessed 2020-08-19]
27. cbitstech/Purple-Robot. URL: https://github.com/cbitstech/Purple-Robot [accessed 2020-08-19]
28. AWARE: Open-source Context Instrumentation Framework For Everyone. URL: https://awareframework.com [accessed 2020-08-19]
29. Aware Framework. URL: https://github.com/awareframework [accessed 2020-08-19]
30. Matcham F, Barattieri di San Pietro C, Bulgari V, de Girolamo G, Dobson R, Eriksson H, et al. Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): a multi-centre prospective cohort study protocol. BMC Psychiatry 2019 Feb 18;19(1):72. [doi: 10.1186/s12888-019-2049-z]
31. RADAR-base. URL: https://github.com/RADAR-base [accessed 2020-08-19]
32. RADAR-CNS. URL: https://www.radar-cns.org [accessed 2020-08-19]
33. Dagum P. Digital biomarkers of cognitive function. npj Digital Med 2018 Mar 28;1(1):00-00. [doi: 10.1038/s41746-018-0018-4]
34. Insel TR. Digital Phenotyping: Technology for a New Science of Behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
35. LeGates TA, Fernandez DC, Hattar S. Light as a central modulator of circadian rhythms, sleep and affect. Nat Rev Neurosci 2014 Jul;15(7):443-454 [FREE Full text] [doi: 10.1038/nrn3743] [Medline: 24917305]
36. Barnett I, Jukka-Pekka O. Inferring mobility measures from GPS traces with missing data. Biostatistics 2020 Apr;21(2):e98-e112. [doi: 10.1093/biostatistics/kxy059]
37. Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time Series Analysis: Forecasting and Control. 5th ed. Hoboken, New Jersey: John Wiley & Sons; 2015.
38. Rasmussen C, Williams C. Gaussian Processes For Machine Learning. Cambridge, Mass: MIT Press; 2008.
39. Ong J, Lau T, Massar S. COVID-19 Related Mobility Reduction: Heterogenous Effects on Sleep and Physical Activity Rhythms. ArXiv Preprint 2006:00-00 (forthcoming).

Abbreviations

API: Application Programming Interface
AWS: Amazon Web Services
CB: circuit breaker
HOPES: Health Outcomes through Positive Engagement and Self-Empowerment
HOPE-S: Health Outcomes via Positive Engagement in Schizophrenia
iOS: iPhone Operating System
QR: Quick Response
R&D: Research and Development

Edited by G Eysenbach; submitted 31.08.20; peer-reviewed by JP Onnela, S Mukherjee; comments to author 10.10.20; revised version received 04.12.20; accepted 18.01.21; published 15.03.21.

Please cite as:
Wang X, Vouk N, Heaukulani C, Buddhika T, Martanto W, Lee J, Morris RJT
HOPES: An Integrative Digital Phenotyping Platform for Data Collection, Monitoring, and Machine Learning
J Med Internet Res 2021;23(3):e23984
URL: https://www.jmir.org/2021/3/e23984
doi:10.2196/23984
PMID:33720028

Viewpoint

Using Narrative Evidence to Convey Health Information on Social Media: The Case of COVID-19

Anat Gesser-Edelsburg[1,2], PhD

[1]School of Public Health, University of Haifa, Haifa, Israel

[2]Health and Risk Communication Research Center, University of Haifa, Haifa, Israel

Corresponding Author:
Anat Gesser-Edelsburg, PhD
School of Public Health
University of Haifa
199 Aba Khoushy Ave
Mount Carmel
Haifa, 3498838
Israel
Phone: 972 544243530
Email: ageser@univ.haifa.ac.il

Abstract

During disease outbreaks or pandemics, policy makers must convey information to the public for informative purposes (eg, morbidity or mortality rates). They must also motivate members of the public to cooperate with the guidelines, specifically by changing their usual behavior. Policy makers have traditionally adopted a didactic and formalistic stance by conveying dry, statistics-based health information to the public. They have not yet considered the alternative of providing health information in the form of narrative evidence, using stories that address both cognitive and emotional aspects. The aim of this viewpoint paper is to introduce policy makers to the advantages of using narrative evidence to provide health information during a disease outbreak or pandemic such as COVID-19. Throughout human history, authorities have tended to employ apocalyptic narratives during disease outbreaks or pandemics. This viewpoint paper proposes an alternative coping narrative that includes the following components: segmentation; barrier reduction; role models; empathy and support; strengthening self-efficacy, community efficacy, and coping tools; preventing stigmatization of at-risk populations; and communicating uncertainty. It also discusses five conditions for using narrative evidence to produce an effective communication campaign on social media: (1) identifying narratives that reveal the needs, personal experiences, and questions of different subgroups to tailor messaging to produce targeted behavioral change; (2) providing separate and distinct treatment of each information unit or theory that arises on social networks; (3) identifying positive deviants who found creative solutions for stress during the COVID-19 crisis not found by other members of the community; (4) creating different stories of coping; and (5) maintaining a dialogue with population subgroups (eg, skeptical and hesitant groups). The paper concludes by proposing criteria for evaluating the effectiveness of a narrative.

(J Med Internet Res 2021;23(3):e24948) doi:10.2196/24948

KEYWORDS

health and risk communication; social media; narrative evidence; crisis; pandemic; misinformation; infodemic; infodemiology; COVID-19; policy; segmentation; barrier reduction; role models; empathy and support; strengthening self/community-efficacy; coping tools; preventing stigmatization; at-risk populations; communicating uncertainty; positive deviance; tailor messaging; targeted behavioral change

Introduction

Background

During disease outbreaks or pandemics, organizations must convey effective information that will cause members of the general public to cooperate with guidelines and even change their behavior, as in the need for social distancing and isolation during the COVID-19 crisis [1]. Moreover, policy makers must convey information to health care professionals, who must deal with new care conditions and social situations [2]. This information must go beyond factual information such as morbidity and mortality statistics. It must also provide explanations to help the public understand the rationale behind the guidelines as well as information to help population

subgroups cope with social conditions such as loneliness and anxiety caused by lifestyle changes.

In communicating this information, health organizations must also address the psychological, sociological, economic, and political factors motivating the behavior of diverse population groups; therefore, conveying information, messages, and guidelines to the public becomes quite complex [3,4]. Moreover, in a media- and communication platform-saturated environment, if policy makers do not convey information that is relevant to people's needs, the public will lose interest and turn to other resources and channels [5-7].

Health information can be conveyed as statistical evidence or as narrative evidence. Statistical evidence usually entails a dry summary of quantitative information about a sample of cases that can be generalized to an entire population [8]. This information is conveyed in a statistics-based and didactic manner that appeals primarily to cognitive considerations.

In contrast, health information can also be conveyed in the form of narrative, through stories that address both cognitive and emotional aspects. Narrative evidence is constructed in the form of a plot that has a beginning, a middle, and an ending that is often open [9-13]. Stories involve characters who portray incidents, life experiences, problems, conflicts, or questions, and challenges emerging from their daily lives or during crises. These characters transcend their personal stories to represent communal stories that often entail information about goals, plans, actions, and outcomes [9-13].

In examining whether narrative or nonnarrative [9,13-15] means are most effective in conveying health care information and creating health behavior change, research has uncovered apparently contradictory results. For example, one study found that narrative evidence is more effective than statistical evidence [16], whereas a meta-analysis indicated that statistical evidence is more persuasive [8].

In another meta-analysis, Zebregs et al [17] identified the influential factors in the two approaches: statistical evidence versus narrative evidence. Statistical evidence was found to exert a stronger influence on beliefs and attitudes than narrative evidence, whereas narrative evidence had a stronger influence on intention. The authors' explanation was that statistical evidence, beliefs, and attitudes are mainly related to cognitive responses, whereas both narrative evidence and behavioral intention are more specifically related to affective (emotional) responses. Accordingly, during a pandemic or other crisis, policy makers can employ both means of information transmission: statistical evidence and narrative evidence. As noted by Zebregs et al [17], narratives can help influence people's intentions to change their behavior, as required by unusual situations.

Hoper and Clippard [18] identified five qualities of narrative messages that make them particularly promising for health interventions. Narrative messages can overcome resistance toward the advocated health behavior. Moreover, they can engage audiences that are less involved, reach audiences that are less knowledgeable, render complex information comprehensible, and ground messages in the culture and experiences of the target audience. In the next section, we describe these five qualities and tie them to the field of emerging infectious disease communication. We also add two qualities that we believe are of particular importance for the use of narrative strategy in health communication. The first is using aesthetic means (nonverbal communication) to convey information based on the edutainment theoretical framework [19,20], as such means are important components of persuasion strategies. The second is conveying a diffused story through social networks based on the diffusion of innovations theory [21,22] and parasocial interactions [23-25], both adapted to the current social media realm.

Overcoming Resistance to Advocated Health Behavior

Resistance can be broadly defined as a reaction against change or an incentive to oppose persuasive appeals. Resistance to persuasive messages may include counterarguing the message's claims, ignoring the message altogether, or denying the validity of the message due to its source. The greater the public's resistance, the greater the advantages of the narrative approach in reaching people [26].

Rendering Complex Information Comprehensible

Narrative evidence helps people process new or complex information by putting the facts into the context of a specific time and place during an outbreak or pandemic. Moreover, narratives can be used to link the information to the experience of the readers or listeners [27,28].

Reaching Audiences That Are Less Knowledgeable

Policy makers seeking to find the most effective way to lead the public to heed information and guidelines during a crisis need a tool that does not require a high level of literacy or education. Narrative evidence meets that requirement because it can address people at all levels and in all languages [18].

Engaging Audiences That Are Less Involved

Addressing emotions is one way to make the public feel involved. Emotions have long been acknowledged as an essential ingredient in the recipe for persuasion [29-31]. In the health communication field, persuasive messages that arouse emotions tend to be perceived as more effective than those that do not [32-36]. Even health care workers may be more responsive to messages that address their emotions than to statistical data. Hence, this approach can be used to provide health care workers with tools and vital information to help them communicate with families whose loved ones are hospitalized with COVID-19.

Grounding Messages in the Culture and Experiences of the Target Audience

People exposed to other people's stories in the media undergo a "parasocial relationship" [37] in that they become engaged with the characters despite never having met them. In line with social-cognitive theory [38], such characters may serve as role models for appropriate behavior by demonstrating the costs and benefits of different courses of action. People may be inspired to imitate the actions of positive characters, avoid the problems of negative characters, or follow in the footsteps of characters who undergo a transformation (usually from negative to positive attitudes or behaviors) over the course of the story [39].

In media campaigns, the characters in a narrative can serve as role models for the readers or viewers [40]. For instance, characters representing at-risk population groups can depict different coping situations that the public can learn and imitate [41]. As opposed to merely short text (eg, a Tweet that can contain up to 280 characters), social media platforms such as Facebook, blogs, or COVID-19 forums enable people to create and share their stories. These platforms provide people with opportunities to talk about their fears and concerns as well as their beliefs and risk perceptions in different situations [5,42,43]. Policy makers can study stories on social networks to learn how people understand the epidemic narrative at any given time and use that information to generate appropriate narratives.

Stories can be used for long-term interventions as well as for the short-term needs of a specific context, such as the outbreak of a pandemic. Furthermore, pandemics are not necessarily short. Indeed, the COVID-19 crisis is prolonged and ongoing. For example, policy makers can design stories that model effective behavior for the subgroup of health care workers caring for older people during the COVID-19 crisis.

Using Aesthetic Means to Convey Information

The newly proposed quality of aesthetic means is of particular importance in health communication. Narratives can use aesthetic means and strategies such as empathy, humor, sarcasm, and irony to convey information using [27]. Edutainment has shown that aesthetic experiences provide viewers with opportunities for meaningful cognitive illumination or change in the context of health or other issues [19,20]. Aesthetic means offer added value that cannot be achieved merely by conveying statistical information.

Conveying a Diffused Story Through Social Networks

Conveying a diffused story [21,22] through social media leads to parasocial interactions [23-25] and generates relationships between people that transcend geographic and linguistic borders. These relationships turn strangers into friends and transform passive audiences into active coparticipants [44,45]. People who hear a good story can be expected to share it with others, initiating a pattern of social proliferation, such that messages "go viral" [46,47]. Thus, the boundaries between the personal and the public become blurred. For example, when people identify with a story posted on social networks and share it on their feeds, they actually turn that story into "their" story. Hence, one individual's story becomes the story of many other individuals, who identify with it and share it with others.

This viewpoint paper combines three intertwining parts to provide a holistic perspective on the use of narrative evidence during a disease outbreak or pandemic. The first part compares the commonly used strategy of apocalyptic narratives to the more desirable strategy of coping narratives, an alternative that has not yet been fully implemented. The second part outlines the conditions necessary to generate an alternative coping narrative and discusses the outcomes of this alternative. Finally, the third part proposes evaluation criteria that can be used in constructing an alternative coping narrative.

Objectives

First Objective

The first objective of this viewpoint is to introduce policy makers to the advantages of using narrative evidence [9,13-15] during a disease outbreak or pandemic, such as the COVID-19 pandemic. To date, health organizations have used narratives mainly in the fields of clinical care and education. These narratives usually focus on disease prevention, disease management, patient recovery, and psychological and social resilience [17]. However, using narrative evidence as a tool for changing attitudes and behaviors can be effective not only for long periods of clinical care intervention but also for shorter periods that require the public to change its behavior.

Second Objective

The second objective of this viewpoint is to propose an alternative coping narrative based on health and risk communication approaches and models. Throughout human history, authorities have tended to employ apocalyptic narratives that include threats, intimidation [48], and the use of "good" and "bad" protagonists. However, alternatives are available to this apocalyptic narrative.

Third Objective

The third objective of this viewpoint is to propose five conditions for constructing and using alternative narrative evidence to launch a communication campaign on social media.

Fourth Objective

The fourth objective of this viewpoint is to propose criteria for evaluating a narrative's effectiveness and potential to generate change: narrative mechanisms, rhetorical concerns, and empirical questions.

Based on the aforesaid, policy makers can use narrative evidence not only for long-term interventions but also during disease outbreaks and pandemics when members of the public are called upon to follow guidelines and change their behavior. In the next section, we propose an alternative coping narrative model instead of the apocalyptic narrative model commonly used during disease outbreaks and pandemics.

Traditional Use of Narrative Evidence in Pandemics: Apocalyptic Narratives

Pandemics are difficult and complex events with a high level of uncertainty. From the dawn of history, pandemics have aroused fear, panic, and alarm, as expressed in many Western works of literature and art [49-54]. Over the years, human and technological progress has led to the development of vaccinations. Nevertheless, epidemics and pandemics, such as the COVID-19 pandemic, still pose a serious challenge, with wide-ranging existential consequences that spark primeval emotions and fears. Questions arise [55-58] such as "How can leaders deal with the public's fears, uncertainty and concerns?" and "What narrative can policy makers create in the public sphere to gain people's trust and cooperation?"

Some, though not all, health organizations currently employ apocalyptic narratives [59]. This sort of narrative lacks many of the qualities of narrative evidence while also containing some elements that can generate negative responses among the public. In this section, we describe the features of apocalyptic narratives traditionally used during pandemics and discuss why these have not been effective. After that, we describe an alternative coping narrative based on the health and risk communication literature that some countries have put into effect during the COVID-19 pandemic.

Throughout human history, pandemic narratives have incorporated melodramatic and apocalyptic features [59,60]. Indeed, the word "epidemic" refers to something that "falls upon people" (in Greek, *epi* means "upon or above" and *demos* means "people") [61]. Hence, by definition, epidemics are unpredictable and are therefore perceived as threatening.

Artistic expressions of epidemics in literature, painting, sculpture, and other media symbolize the sense of vulnerability in the face of uncertainty and death, as well as the random nature of death itself. The villain of the plot is the virus that is threatening to destroy humanity, while the "good guys" or heroes are the lifesaving medical workers. The narrative also includes characters depicted as disease spreaders, usually from disempowered communities. Blame, stigmatization [62], and fears and anxieties (whether real or exaggerated) swirl around in the public consciousness [62]. The tone of this narrative is apocalyptic rather than redemptive. Diseases are managed and endured rather than overcome, and species-level damage is incurred. In pandemic narratives, our anxieties are not assuaged; we are invited to struggle rather than to overcome.

According to Wald, pandemic narratives tell "a contradictory but compelling story of the perils of human interdependence and the triumph of human connection and cooperation, scientific authority, and the evolutionary advances of the microbe, ecological balance, and impending disaster" [62]. Further, Massumi [63] indicated that we live in an environment that is not so much threatening as "threat generating" [64]. That is, the threat is not always as existential as its effect on human consciousness, as expressed through the stories we tell.

In the modern pandemic narrative, traditional and social media do not only cover and mediate the crisis; they also serve as narrators that dictate the reality and narrative of the pandemic to the audience. In this narrative, humanity searches for a solution in the form of a medication or vaccination that will redeem it from the apocalyptic threats [3]. In recent disease outbreaks, health organizations seem to have strengthened this apocalyptic narrative by using strategies of intimidation to make the public follow instructions and guidelines [3]. This can be seen in the language and tone of information delivery (eg, use of war language to describe COVID-19 as a cruel enemy that needs to be defeated [65]).

Moreover, the modern pandemic narrative often uses overblown statistics not backed by accurate facts to describe morbidity and mortality to motivate the public to follow directives. For example, Dew [66] describes how during the 1997 measles outbreak in New Zealand, the Ministry of Health ran a media advertisement campaign using emotional appeals and employing statistics and numbers to create a "quantification rhetoric." According to Petersen and Lupton [67], this rhetoric "tends to suggest the figures used are not subject to doubt or uncertainty." During the media campaign, "the viewer was subjected to images of cemeteries and crucifixes passing across the screen" [66]. However, the 1997 outbreak in New Zealand was found to be minor. The actual number of measles cases reported was 1200, and not a single child died [66]. Intimidation has also been used during the COVID-19 pandemic. For example, the prime minister of Israel compared the first wave of the epidemic to both Spanish influenza and the Holocaust, citing inaccurate statistics [68,69].

Policy makers and organizations often tend to frame uncertain information in terms of certainty. Their assumption is that uncertain information may create negative emotions. Furthermore, even when the risk is uncovered, often through social media, and its communication becomes inevitable, experts and organizations are often reluctant to reveal all available information. They prefer to provide a straightforward and unambiguous explanation. Van Asselt et al [70,71] called this framing "the uncertainty paradox," referring to situations wherein uncertainty is acknowledged, but the role of science is framed as providing certainty [72].

Contrary to this assumption, other studies indicate that when people feel they do not have sufficient information regarding a risk, their sense of uncertainty and negative feelings may increase [73-76], especially when the risk is perceived as severe and uncontrollable [77]. Indeed, honest risk communication and providing sufficient information do not have a negative impact on the public's behavior. In contrast, sufficient and accurate information can help mitigate negative feelings [78-81].

Authorities often use intimidation strategy because they believe the public is in a state of "panic" and "hysteria" during a crisis [82]. For example, the public's reaction to the appearance of four Ebola cases in the United States and to the authorities' diverse approaches to necessary precautionary measures was perceived as "national panic" [83,84], with Maryn McKenna [85] coining the term "Ebolanoia" to describe it.

Contrary to this widely accepted view of public panic, empirical studies of public response to extreme situations have revealed the opposite findings [79,80,86]. Indeed, some studies indicate that in extreme situations, people are more likely to react by demonstrating social cohesion and mutual trust rather than showing panic [87].

Even in the case of public panic, using intimidation without empowering individual self-efficacy is counter to the theory of intimidation use known as the extended parallel process model (EPPM) [88]. The EPPM attempts to predict how individuals will react when confronted with fear-inducing stimuli. For fear-based policies to be effective, policy makers must induce a moderate level of fear alongside a higher level of self-efficacy and response efficacy. When the public's fear exceeds its sense of self-efficacy, the message becomes ineffective.

An Alternative Coping Narrative

As opposed to this apocalyptic narrative, here, we propose an alternative coping narrative based on health and risk communication approaches and frameworks [89,90]. This narrative should contain the following components: segmentation, barrier reduction, role modeling, empathy and support, tools to promote self and collective efficacy and coping, preventing the stigmatization of at-risk populations, and communication of uncertainty.

Segmentation Through Narrative

The literature underscores the importance of segmenting [91,92] and mapping [93,94] each subgroup in the population to tailor [95,96] the information and media campaign to the barriers, risks, concerns, and unique needs of each group. During every disease outbreak or pandemic, some groups are at higher risk than others. The narrative put forward by the authorities must communicate and distinguish between actions taken for the benefit of the public at large and those targeting specific at-risk groups. For example, during the COVID-19 pandemic, young people between the ages of 18 and 30 years without any underlying conditions are at lower risk of serious illness. Therefore, the authorities must tailor risk messages to particular at-risk groups without resorting to intimidation.

Reducing Barriers Through Narrative

The strategy of barrier reduction entails reducing existing difficulties and barriers to the adoption of desirable behavior [97,98] and offering incentives and solutions to the population. This strategy can be useful during a disease outbreak or pandemic. One of the barriers to adoption of desirable behavior during the COVID-19 pandemic is the difficulty of maintaining social distancing. By means of narratives that illustrate this barrier while providing ways of coping with it, the public can be given solutions for complying with social distancing without the use of intimidation.

Role Modeling Through Narrative

Research has shown that role models, identification, and social support can be used effectively in interventions to change health behavior [38,99]. According to Bandura's sociocognitive theory, individuals can learn a behavior by observing a model. Moreover, they will be more likely to perform this behavior if they see positive and appealing reinforcement for the behavior. The use of role models boosts self-efficacy in that the characters demonstrating a particular health behavior provide viewers with tools and skills.

The use of role models to teach social behavior through narratives can be implemented during disease outbreaks or pandemics as well. During the COVID-19 pandemic, for instance, narratives using positive role models can demonstrate the advantages of following the guidelines, thus strengthening people's self-efficacy. Likewise, patients who survived COVID-19 can share their experiences and give tips to the public. Leaders dealing with the crisis can also serve as role models through their behavior. For example, during the COVID-19 crisis, New Zealand's Prime Minister Jacinda Ardern announced a 20% salary cut for herself and the members of her cabinet [100].

Strengthening Collective Efficacy Through Narrative

Beyond strengthening individual self-efficacy, narratives can strengthen collective efficacy by illustrating the community's ability to provide social support for its members. A community's collective efficacy can be reinforced through stories that emphasize solidarity and mutual support for weaker community members during a health crisis or pandemic. For example, during the COVID-19 pandemic, civic organizations and individuals can support older people under lockdown by helping them obtain food and medicine.

Using Narratives to Prevent Stigmatization of At-Risk Populations

During a health crisis, authorities sometimes worry that at-risk population groups will reject relevant information for fear of being stigmatized by the media and society. The literature points to the possibility of self-stigmatization or social stigmatization if media outlets use sensational means to communicate a risk [101]. For example, during the COVID-19 pandemic, the Asian American community expressed strong fears of being blamed for the spread of SARS-CoV-2. Therefore, policy makers should stress stigma reduction and create narratives that underscore social solidarity.

Communicating Uncertainty Through Narrative

Scholars investigating the topic of risk found that dealing with uncertainty is a major challenge in risk assessment and management. According to Frewer et al [102], public health experts tend to believe that the public is incapable of coping with the uncertainties associated with risk management. Contrary to this opinion, studies in the risk communication literature indicate that in risky situations [70,71,103,104], especially those that involve uncertainty [105], the public wants full information transparency. Transparent communication does not provoke negative reactions among the public; rather, it helps reduce negative feelings and increases the public's respect for the risk-assessing agency [79].

Sandman and Lanard [106] emphasize the need to "proclaim uncertainty," advising authorities to disseminate tentative information if it is the only type of information they have. A number of studies conducted during pandemics, such as the Ebola outbreak in the United States [107] and the polio outbreak in Israel [108], reinforced Sandman and Lenard's hypothesis by showing that the public wants organizations to communicate uncertainty. Furthermore, the public becomes impatient and uncooperative when authorities only give them partial or selective information [109].

Table 1 summarizes the strategies and components of a pandemic coping narrative based on health and risk communication approaches. For each apocalyptic pandemic narrative, an alternative pandemic narrative that offers coping strategies is presented to help health organizations transform one narrative to the other. Figure 1 depicts an apocalyptic pandemic narrative, in which COVID-19 is depicted as an apocalyptic explosion of an atomic bomb causing severe harm

to humans. Figure 2 depicts an alternative pandemic narrative that offers coping strategies, using additional tools provided to people to cope with the COVID-19 pandemic.

In this section, we explained how an apocalyptic narrative can be transformed into a coping narrative. In the next section, we propose several conditions necessary for constructing and using a coping narrative to launch a communication campaign on social media.

Table 1. Apocalyptic pandemic narratives versus alternative coping narratives.

Apocalyptic pandemic narrative	Alternative pandemic narrative that offers coping strategies
Waging war against an enemy	Coping with situations of uncertainty
Using intimidation strategies to motivate the public to follow guidelines	Using empathy strategies and reflexivity to motivate the public to cooperate
Creating heroes (leaders/life-saving medical teams)	Creating social support and mutual aid through health organizations
Prioritizing public health as the most important thing	Differentiating between public health and personal risk
Taking extreme measures to protect the public	Introducing fact-based measures
Using sensationalism and dramatization	Seeking truth and exposing policy makers' doubts and questions
Enacting surveillance, guidelines, and regulations ("Big Brother")	Transparency and rationalization of guidelines
Stigmatizing and blaming groups that do not follow guidelines	Encouraging solidarity
Closed ending: defeat or victory over the virus	Coping and living in a changing and dynamic situation

Figure 1. An apocalyptic pandemic narrative.

Figure 2. An alternative pandemic narrative that presents coping strategies.

Use of Narrative Evidence to Communicate About COVID-19 via Social Media While Maintaining Constructive Dialogue With the Public

A narrative media campaign launched on social media can be based on one or both of the following methods: (1) posting personal stories on social networks and distributing them to relevant subgroups in the population via channels targeting these groups; (2) using narratives based on preliminary research that identifies the public's questions and concerns and responds to them through narrative evidence posted on social networks. Each of these methods requires five main conditions. In the following section, we outline these conditions, methods, and prospective outcomes. These conditions and their outcomes are formulated based on integrating health and risk communication theories [3,4,110]. We thus provide a new perspective on the use of narrative and communication strategies during disease outbreaks and pandemics—in this case, the COVID-19 pandemic. These conditions and outcomes have a high degree of reliability and can be further validated by additional empirical research.

First Condition: Tailor Messages Toward Targeted Behavioral Change Based on the Needs and Experiences of Different Subgroups

Despite the theoretical understanding that national health authorities should build segmented profiles of their publics [111], this understanding has not yet been fully implemented. During the midst of the H1N1 pandemic, countries were called on to adapt their communication strategies to specific cultural needs [112], pointing to a general lack of such cultural and social adaptation [111]. Although government agencies have long recognized the ineffectiveness of one-size-fits-all messaging [113], studies have indicated that segmentation is still far from adequate implementation [114-118].

Method

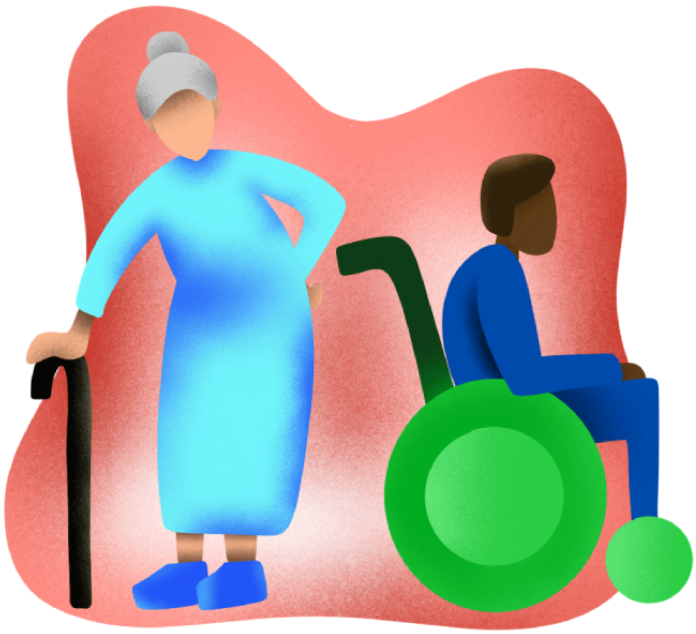
Content analysis [119] and ethnographic analysis [120] should be used to map and categorize the narratives of specific subgroups on social media. The variables defining such groups will vary depending on the issue. In addition to sociodemographic or geodemographic variables typifying different countries, other variables will be based on specific attributes of different groups (eg, trust in authorities, science skepticism, and vaccination hesitancy). Through this research apparatus, policy makers can use qualitative and quantitative tools to map and analyze the stories arising from different population groups and the theories to be elicited from them.

Outcomes

Health organizations will be able to understand the needs, needs, assumptions, and risk perceptions of different groups and respond instantly. When health organizations identify the main stories of each subgroup, they will be able to adjust the relevant information accordingly. Figure 3 depicts the need to identify the main stories of each subgroup, such as older or disabled people.

Health organizations can use people's authentic stories to disseminate essential information to the community. When health organizations use the experiences of people who found ways to cope with different crisis situations, other people can learn from that information, thus building social resilience.

Figure 3. Identifying the main stories of each subgroup.



Second Condition: Refer Separately and Distinctly to Each Information Unit or Theory Arising on Social Networks

Studies show that when health organizations want to communicate facts to the public, they often distinguish between myth and fact [121-123]. This distinction is not neutral and has been found to be ineffective for two reasons. First, when information provided on a website is identified as a myth, people still remember the information, even though it is totally or partially untrue. Second, the public refuses to accept a judgmental approach without scientific evidence. In two studies on public attitudes toward the measles-mumps-rubella vaccine and the seasonal influenza vaccine [124,125], pro-vaccine information from the US Centers for Disease Control website had a "backfire effect." After being given information intended to refute the supposed connection between vaccinations and autism, vaccine skeptics formed even stronger negative opinions about vaccinations.

Health organizations must provide separate and distinct treatment for any kind of information unit or theory that arises on social networks. For example, social media platforms are filled with rumors pertaining to COVID-19 [126,127]. Health organizations have generally used a single approach to handle information they consider unfounded, without sufficient differentiation. Thus, they countered the claim that the virus originated in a laboratory in Wuhan in the same way they countered the proposal to eat garlic as a cure or the notion that the virus can be killed through exposure to sunlight. These units of information differ, and each deserves to be engaged and addressed on its own merits.

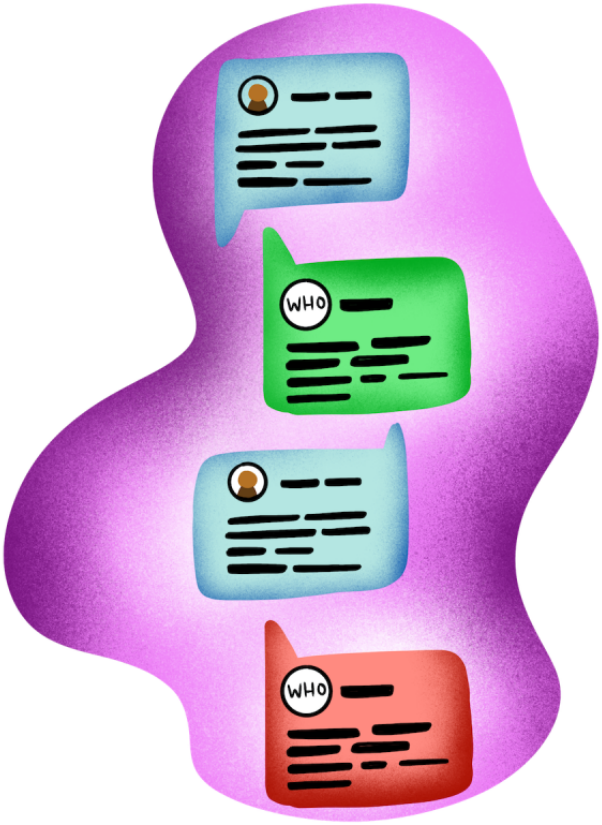
Method

Answer the questions and theories posed by different population groups, not by correcting the information but rather by differentiating the information and addressing each claim on its own merits.

Outcome

By distinguishing among different theories that arise on social media and addressing them separately, health authorities will build the public's trust. Health and risk communication theories show that bidirectional dialogue is critical, that is, the positions and arguments emerging from a theory should be addressed through a dialogue between equals. Likewise, conveying positive feedback regarding the factual parts of different theories raised on social networks will give the public a sense of transparency and trust. In contrast, deciding to correct or dismiss entire theories (including their correct parts) can generate antagonism, such that the public feels its views are being dismissed. Figure 4 depicts the need to conduct a dialogue between health organizations and the public regarding COVID-19–related concerns and questions raised by the public.

Figure 4. Answering the questions and theories posed by different population groups.



Third Condition: Identify Positive Deviants That Offer Creative Solutions

According to Singhal [128], "the Positive Deviance (PD) approach is based on the premise that in every community there are certain individuals or groups whose uncommon behaviors and strategies enable them to find better solutions to problems than their peers, while facing worse challenges and having access to the same resources. However, these people are ordinarily invisible to others in the community." The PD approach seeks to identify and streamline existing resources deriving from within a community rather than to import external "best practices." Such practices are distributed and implemented over time via social networks [129,130].

Health organizations should seek out positive deviants [131-133] who propose creative ("outside the box") solutions for stressful situations emerging from the COVID-19 crisis that other members of the community did not find.

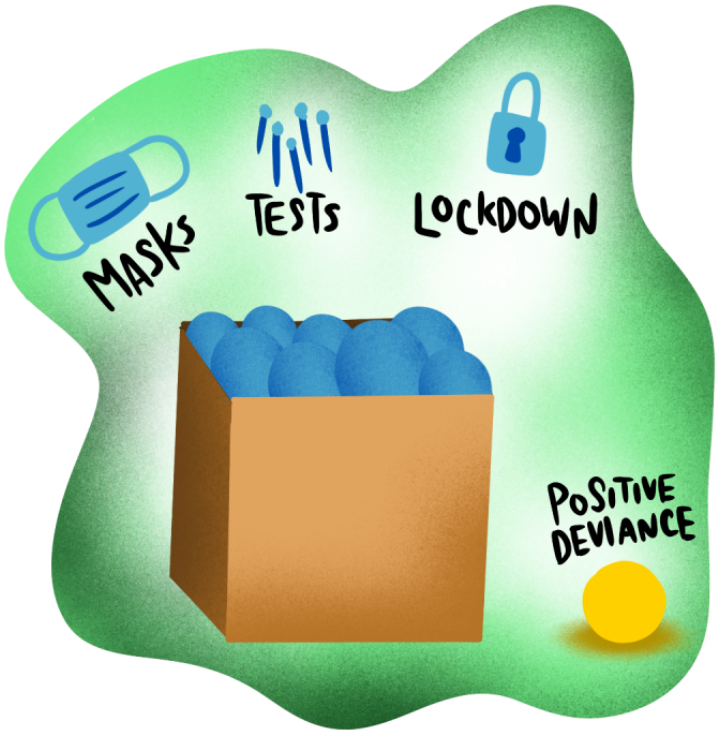
Method

Health organizations should use the narratives of exceptional individuals in various groups who have found ways to cope with loneliness, stress, and pressure. These coping means can then be disseminated to other members of their community.

Outcome

These creative solutions and "thinking outside the box" will generate role models and promote tips from ordinary people representing various population groups that can help the public cope during the COVID-19 crisis. The advantage is that the community is more likely to accept solutions coming from inside than those imposed by the authorities. Figure 5 visualizes the need to think outside the box to find creative solutions that can be adapted to the changing state of the COVID-19 pandemic.

Figure 5. Thinking outside the box.



Fourth Condition: Create Different Stories of Coping Experiences

Storytelling relies upon realism, identification, and transportation to help people understand different points of view and change their attitudes and health behavior [134]. According to Lee et al [134], "narrative communication is context-dependent because it derives meaning from the surrounding situation and provides situation-based stories that are a pathway to processing story content."

Method

Instead of the dry statistics and didactic guidelines that health authorities convey, members of different subgroups can share their stories with their friends and introduce dilemmas and emotions emerging from their coping experiences. Figure 6 visualizes the need to create different coping stories using strategies such as identification and humor.

These narratives may be in the form of testimonials. They can also be dramatizations of personal narratives [13] that illustrate what happened to the narrator or to other individuals during the crisis (eg, a story about how a patient from an immigrant group copes with stress).

Figure 6. Creating different stories of coping experiences.

Outcome

These stories can provide specific tools to help different population subgroups cope with the crisis.

Fifth Condition: Maintain a Dialogue With Skeptical and Hesitant Groups

According to Larson [135], "educational materials and resources are important, but limited; health officials and educational campaigns often fall short because they craft messages based on what they want to promote, without addressing existing perceptions. Dialogue matters. Strategies must include listening and engagement. We have to get better at this…"

Method

Health authorities can use the authentic narratives and social media posts of skeptical and hesitant groups to answer questions and address arguments while providing objective and transparent information. In doing so, authorities should not attempt to frame the arguments of these hesitant and skeptical groups in terms of myths versus facts or as misinformation.

Outcomes

Building trust among skeptical groups will have consequences for enlisting the cooperation of these groups in future pandemics. Figure 7 depicts the need to maintain a dialogue with skeptical and hesitant groups.

After outlining the conditions underlying the use of narrative evidence to communicate crises, we now propose criteria for evaluating the effectiveness of a narrative.

Figure 7. Maintaining a dialogue with skeptical and hesitant groups.

A Formative Evaluation Toolkit for Health Organizations

Formative evaluation of a narrative must take into consideration both the narrative created by the organization itself and the authentic narratives found on social networks and used by the organization during campaigns. The purpose of formative evaluation is to ensure that the intervention element is applicable, suitable, significant, and acceptable to the program's target audience [136]. Formative evaluation focuses on participatory research with the target audience before, during, and after launching the communication campaign. It includes checking the barriers, needs, and preferences of the target audiences and setting objectives on the way to designing the narrative. Formative evaluation for narrative building should be preceded by qualitative empirical research among representatives of the target audience (including personal interviews, focus groups, and role playing).

Table 2 summarizes the questions and issues relevant to examining a narrative through evaluation research.

Table 2. A rhetorical matrix for empirical analysis of narrative mechanisms and potential for change (evaluation toolkit), based on Gesser-Edelsburg and Singhal (2013) [19].

Narrative mechanism	Rhetorical concerns	Empirical questions to gauge a narrative's potential for change
Dialogue (between the narrative and the public)	How do the produced messages and dialogue engage with the public's predisposed realities?	In processing the narrative, to what extent did the public feel • They were invited or coerced into a dialogue about coping with the challenges? • The messages were consensual or oppositional to their predispositions? • New possibilities for coping were raised in the narrative?
Involvement (the public's emotional engagement with the narrative)	How is the public emotionally involved, immersed, or absorbed in the unfolding narrative?	In processing the narrative, to what extent did the public experience • Feelings of voyeurism, empathic identification, alienation, or anger? • Identification with certain characters, and how did that influence their perceptions and positions on the issues the characters represented?
Trust (public's perceptions of the narrative's credibility)	How does the public perceive the plausibility, realism, and veracity of the unfolding narrative? Is the narrative trustworthy? Credible?	• In processing the narrative, to what extent did the public feel the narrative was credible? Realistic? Plausible? • At what stage did the public begin to experience clarification of doubts and new emergent possibilities? What conditions facilitated this change?
Catharsis and transformation (narrative's influence on the public)	How does public engagement with the narrative lead to new learning, alternative positions, and change possibilities? How does the modeling and reinforcement of change through characters increase audience motivation and self-efficacy for practice?	In processing the narrative, to what extent did the public feel • They identified with the transformation of characters in the unfolding story? • They went through a process of change parallel to the transformed characters? • They were engaged and empowered by the characters and their story? • The alternatives presented in the narratives are applicable to the reality of their behavior?

Conclusions

The use of narrative evidence as a tool for changing attitudes and behaviors is effective not only for long periods of clinical care intervention but also for short ones, because in either case, the public is required to change its behavior. As we have realized during the COVID-19 pandemic, the public will be forced to change its lifestyle over the long term.

During a disease outbreak or pandemic, policy makers must deal with the flow of information on multiple media forums. Indeed, policy makers must compete for the public's attention with other sources that may be manufacturing misinformation. In such a complex multimedia environment, the use of narrative has many advantages.

Seven qualities of narrative messages make them particularly promising for health interventions. Narrative messages can overcome resistance toward the advocated health behavior, engage audiences that are less involved, reach audiences with less knowledge, render complex information comprehensible, ground messages in the target audience's culture and experience [18], use aesthetic means, and convey a diffused story over social networks.

Throughout human history, authorities have tended to employ apocalyptic narratives during disease outbreaks or pandemics. This viewpoint paper proposes an alternative coping narrative model based on health and risk communication approaches and models incorporating the following components: segmentation [137]; barrier reduction [97,98]; role models, empathy, and support [90,99]; strengthening self-efficacy, community

efficacy, and coping tools [89]; preventing the stigmatization of at-risk populations; and communicating uncertainty.

In this viewpoint paper, we also recommend five conditions for using narrative evidence that will lead to launching an effective communication campaign on social media:

1. Identifying narratives that reveal the needs, personal experiences, and questions of different groups to tailor messaging toward producing targeted behavioral change
2. Offering separate and distinct treatment of each information unit or theory of any kind that arises on social networks
3. Identifying positive deviants [131-133] who have found creative solutions for stress during the COVID-19 crisis that other members of the community did not find
4. Creating different stories of coping experiences
5. Maintaining a dialogue with subgroups (eg, skeptical and hesitant groups)

Evaluating the narrative constructed by health organizations is also very important. In this viewpoint paper, we offer criteria for evaluating the effectiveness of a narrative by addressing narrative mechanisms**,** rhetorical concerns, and empirical questions to gauge each narrative's potential for change.

The proposed use of narrative as a communication tool will help policy makers more effectively manage how they communicate with the public during disease outbreaks and pandemics. Narrative is a human and pluralistic means that appeals to everyone. Hence, by using existing narratives on social networks while simultaneously creating new narratives to transmit information, health officials and policy makers are more likely to be able to influence actual health attitudes and behaviors.

Conflicts of Interest

None declared.

References

1. Social Distancing: keep a safe distance to slow the spread. US Centers for Disease Control and Prevention. 2020 Nov 06. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html [accessed 2021-03-10]
2. Gesser-Edelsburg A, Stolero N, Mordini E, Billingsley M, James JJ, Green MS. Emerging infectious disease (EID) communication during the 2009 H1N1 influenza outbreak: literature review (2009-2013) of the methodology used for EID communication analysis. Disaster Med Public Health Prep 2015 Apr;9(2):199-206. [doi: 10.1017/dmp.2014.126] [Medline: 25882126]
3. Gesser-Edelsburg A, Shir-Raz Y. Risk Communication and Infectious Diseases in an Age of Digital Media. London, UK: Routledge; 2016.
4. Lundgren R, McMakin A. Risk Communication: A Handbook for Communicating Environmental, Safety, and Health Risks. 5th edition. Piscataway, NJ: IEEE Press; 2013.
5. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
6. Hughes A, Palen L, Sutton J, Liu S, Vieweg S. "Site-seeing" in disaster: an examination of on-line social convergence. In: Proceedings of the 5th International ISCRAM Conference. 2008 May Presented at: 5th International ISCRAM Conference; May 2008; Washington, DC URL: https://cmci.colorado.edu/~palen/Papers/iscram08/OnlineConvergenceISCRAM08.pdf
7. Stiegler R, Tilley S, Parveen T. Finding family and friends in the aftermath of a disaster using federated queries on social networks and websites. 2011 Presented at: 13th IEEE International Symposium on Web Systems Evolution (WSE); September 30, 2011; Williamsburg, VA p. 21-26. [doi: 10.1109/wse.2011.6081815]
8. Allen M, Preiss RW. Comparing the persuasiveness of narrative and statistical evidence using meta‐analysis. Communication Research Reports 1997 Mar;14(2):125-131. [doi: 10.1080/08824099709388654]
9. Dahlstrom MF. Using narratives and storytelling to communicate science with nonexpert audiences. Proc Natl Acad Sci U S A 2014 Sep 16;111 Suppl 4:13614-13620 [FREE Full text] [doi: 10.1073/pnas.1320645111] [Medline: 25225368]
10. Hinyard LJ, Kreuter MW. Using narrative communication as a tool for health behavior change: a conceptual, theoretical, and empirical overview. Health Educ Behav 2007 Oct;34(5):777-792. [doi: 10.1177/1090198106291963] [Medline: 17200094]
11. Kopfman JE, Smith SW, Ah Yun JK, Hodges A. Affective and cognitive reactions to narrative versus statistical evidence organ donation messages. J Appl Commun Res 1998 Aug;26(3):279-300. [doi: 10.1080/00909889809365508]
12. Larkey LK, Hecht M. A model of effects of narrative as culture-centric health promotion. J Health Commun 2010 Mar;15(2):114-135. [doi: 10.1080/10810730903528017] [Medline: 20390982]
13. Miller-Day M, Hecht ML. Narrative means to preventative ends: a narrative engagement framework for designing prevention interventions. Health Commun 2013;28(7):657-670 [FREE Full text] [doi: 10.1080/10410236.2012.762861] [Medline: 23980613]
14. Murphy ST, Frank LB, Chatterjee JS, Moran MB, Zhao N, Amezola de Herrera P, et al. Comparing the Relative Efficacy of Narrative vs Nonnarrative Health Messages in Reducing Health Disparities Using a Randomized Trial. Am J Public Health 2015 Oct;105(10):2117-2123. [doi: 10.2105/AJPH.2014.302332] [Medline: 25905845]

15. Stavrositu CD, Kim J. All blogs are not created equal: the role of narrative formats and user-generated comments in health prevention. Health Commun 2015;30(5):485-495. [doi: 10.1080/10410236.2013.867296] [Medline: 24978046]
16. Baesler EJ, Burgoon JK. The Temporal Effects of Story and Statistical Evidence on Belief Change. Commun Res 2016 Jun 30;21(5):582-602. [doi: 10.1177/009365094021005002]
17. Zebregs S, van den Putte B, Neijens P, de Graaf A. The differential impact of statistical and narrative evidence on beliefs, attitude, and intention: a meta-analysis. Health Commun 2015 May 16;30(3):282-289. [doi: 10.1080/10410236.2013.842528] [Medline: 24836931]
18. Hopfer S, Clippard JR. College women's HPV vaccine decision narratives. Qual Health Res 2011 Feb;21(2):262-277. [doi: 10.1177/1049732310383868] [Medline: 20841433]
19. Gesser-Edelsburg A, Singhal A. Enhancing the persuasive influence of entertainment-education events: rhetorical and aesthetic strategies for constructing narratives. Crit Arts 2013 Feb;27(1):56-74. [doi: 10.1080/02560046.2013.766973]
20. Sood S, Menard T, Witte K. The theory behind entertainment-education. In: Singhal A, Cody MJ, editors. Entertainment-Education and Social Change: History, Research, and Practice. Mahwah, NJ: Lawrence Erlbaum Associates; 2004:117-149.
21. Rogers E. Diffusion of Innovations. 5th Edition. New York, NY: Free Press; 2003.
22. Rogers EM. A prospective and retrospective look at the diffusion model. J Health Commun 2004 Jan;9 Suppl 1(sup1):13-19. [doi: 10.1080/10810730490271449] [Medline: 14960401]
23. Horton D, Richard Wohl R. Mass communication and para-social interaction. Psychiatry 2016 Nov 08;19(3):215-229. [doi: 10.1080/00332747.1956.11023049] [Medline: 13359569]
24. Rubin AM, Rubin RB. Interface of personal and mediated communication: A research agenda. Crit Stud Mass Commun 2009 May 18;2(1):36-53. [doi: 10.1080/15295038509360060]
25. de Graaf A, Hoeken H, Sanders J, Beentjes JWJ. Identification as a mechanism of narrative persuasion. Commun Res 2011 May 16;39(6):802-823. [doi: 10.1177/0093650211408594]
26. Kreuter MW, Green MC, Cappella JN, Slater MD, Wise ME, Storey D, et al. Narrative communication in cancer prevention and control: a framework to guide research and application. Ann Behav Med 2007 Jun;33(3):221-235. [doi: 10.1007/BF02879904] [Medline: 17600449]
27. Gesser-Edelsburg A. Drama as a rhetorically effective health communication strategy: risks and benefits. In: Kim DK, Singhal A, Kreps GL, editors. Global Health Communication Strategies in the 21st Century: Design, Implementation, and Evaluation. New York, NY: Peter Lang Publishing Group; 2014:134-150.
28. Gray JB. The power of storytelling: using narrative in the healthcare context. J Commun Healthc 2013 Jul 18;2(3):258-273. [doi: 10.1179/cih.2009.2.3.258]
29. Eagly A, Chaiken S. The Psychology of Attitudes. New York, NY: Academic Press; 1993.
30. McGuire WJ. The nature of attitudes and attitude change. In: Lindzey G, Aronson E, editors. The Handbook of Social Psychology. Reading, MA: Addison-Wesley; Apr 1969:181-314.
31. Petty R, DeSteno D, Rucker D. The role of affect in attitude change. In: Forgas JP, editor. Handbook of Affect and Social Cognition. Mahwah, NJ: Lawrence Erlbaum Associates; 2001:212-233.
32. Biener L. Adult and youth response to the Massachusetts anti-tobacco television campaign. J Public Health Manag Pract 2000 May;6(3):40-44. [doi: 10.1097/00124784-200006030-00007] [Medline: 10848481]
33. Biener L, Reimer RL, Wakefield M, Szczypka G, Rigotti NA, Connolly G. Impact of smoking cessation aids and mass media among recent quitters. Am J Prev Med 2006 Mar;30(3):217-224. [doi: 10.1016/j.amepre.2005.10.026] [Medline: 16476637]
34. Dillard J, Nabi R. The persuasive influence of emotion in cancer prevention and detection messages. J Commun 2006;56(s1):S123-SS39. [doi: 10.1111/j.1460-2466.2006.00286.x]
35. Dillard JP, Peck E. Affect and persuasion. Commun Res 2016 Jun 30;27(4):461-495. [doi: 10.1177/009365000027004003]
36. Pechmann C, Reibling ET. Antismoking advertisements for youths: an independent evaluation of health, counter-industry, and industry approaches. Am J Public Health 2006 May;96(5):906-913. [doi: 10.2105/AJPH.2004.057273] [Medline: 16571709]
37. Meyrowitz J. No Sense of Place. New York, NY: Oxford University Press; 1985.
38. Bandura A. Social Learning Theory. Englewood Cliffs, NJ: Prentice Hall; 1977.
39. Rogers EM, Vaughan PW, Swalehe RM, Rao N, Svenkerud P, Sood S. Effects of an entertainment-education radio soap opera on family planning behavior in Tanzania. Stud Fam Plann 1999 Sep;30(3):193-211. [doi: 10.1111/j.1728-4465.1999.00193.x] [Medline: 10546311]
40. Frank LB, Murphy ST, Chatterjee JS, Moran MB, Baezconde-Garbanati L. Telling stories, saving lives: creating narrative health messages. Health Commun 2015;30(2):154-163 [FREE Full text] [doi: 10.1080/10410236.2014.974126] [Medline: 25470440]
41. Wang H, Singhal A. East Los High: transmedia edutainment to promote the sexual and reproductive health of young Latina/o Americans. Am J Public Health 2016 Jun;106(6):1002-1010. [doi: 10.2105/AJPH.2016.303072] [Medline: 27077336]

42. Zhang EX, Yang Y, Shang RD, Simons JJP, Quek BK, Yin XF, et al. Leveraging social networking sites for disease surveillance and public sensing: the case of the 2013 avian influenza A(H7N9) outbreak in China. WPSAR 2015 Apr 06;6(2):66-72. [doi: 10.5365/wpsar.2015.6.1.013]
43. Henrich N, Holmes B. Communicating during a pandemic: information the public wants about the disease and new vaccines and drugs. Health Promot Pract 2011 Jul 08;12(4):610-619. [doi: 10.1177/1524839910363536] [Medline: 21059872]
44. Booth P. Digital Fandom: New Media Studies. New York, NY: Peter Lang Inc; 2010.
45. Baym N. Personal Connections in the Digital Age. Cambridge, UK: Polity Press; 2010.
46. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
47. Kass-Hout TA, Alhinnawi H. Social media in public health. Br Med Bull 2013;108:5-24. [doi: 10.1093/bmb/ldt028] [Medline: 24103335]
48. Doshi P. Are US flu death figures more PR than science? BMJ 2005 Dec 08;331(7529):1412. [doi: 10.1136/bmj.331.7529.1412]
49. Leviton R. A shot in the dark. Yoga Journal 1992;104:59-61.
50. Kloss B, Bruce T. Graphic Guide to Infectious Disease. Philadelphia, PA: Elsevier; 2018.
51. Mitchell P. Contagious Metaphor. London: Bloomsbury Academic; 2012.
52. Moore J. Deadly diseases: epidemics throughout history. CNN. 2014. URL: https://www.cnn.com/interactive/2014/10/health/epidemics-through-history/ [accessed 2021-03-10]
53. Potter P. Art in Science: Selections from Emerging Infectious Diseases. New York, NY: Oxford University Press; 2014.
54. Vidal P, Tibayrenc M, Gonzalez J. Infectious disease and arts. In: Tibayrenc M, editor. Encyclopedia of Infectious Diseases: Modern Methodologies. Hoboken, NJ: Wiley; 2007.
55. Rufai SR, Bunce C. World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis. J Public Health (Oxf) 2020 Aug 18;42(3):510-516 [FREE Full text] [doi: 10.1093/pubmed/fdaa049] [Medline: 32309854]
56. Gesser-Edelsburg A, Cohen R, Hijazi R, Abed Elhadi Shahbari N. Analysis of public perception of the Israeli government's early emergency instructions regarding COVID-19: online survey study. J Med Internet Res 2020 May 15;22(5):e19370 [FREE Full text] [doi: 10.2196/19370] [Medline: 32392172]
57. Ladkin D. What Donald Trump's response to COVID-19 teaches us: it's time for our romance with leaders to end. Leadership 2020 May 21;16(3):273-278. [doi: 10.1177/1742715020929134]
58. Garrett L. COVID-19: the medium is the message. Lancet 2020 Mar;395(10228):942-943. [doi: 10.1016/s0140-6736(20)30600-0]
59. Davis M. "Is it going to be real?" Narrative and media on a pandemic. Forum Qual Soc Res 2017;18(1). [doi: 10.17169/fqs-18.1.2768]
60. Wonser R, Boyns D. Between the Living and Undead: How Zombie Cinema Reflects the Social Construction of Risk, the Anxious Self, and Disease Pandemic. Sociol Q 2016 Nov 16;57(4):628-653. [doi: 10.1111/tsq.12150]
61. Cunningham A. Epidemics, Pandemics, and the Doomsday Scenario. Hist Speak 2008;9(7):29-31. [doi: 10.1353/hsp.2008.0035]
62. Wald P. Contagious: Cultures, Carriers, and the Outbreak Narrative. Durham, NC: Duke University Press; 2008:978.
63. Massumi B. Potential politics and the primacy of preemption. Theory & Event 2007;10(2). [doi: 10.1353/tae.2007.0066]
64. Mitchell S, Hamilton SN. Playing at apocalypse: reading Plague Inc. in pandemic culture. Convergence 2017 Jan 17;24(6):587-606. [doi: 10.1177/1354856516687235]
65. Qian X, Ren R, Wang Y, Guo Y, Fang J, Wu Z, Members of Steering Committee, Society of Global Health, Chinese Preventive Medicine Association. Fighting against the common enemy of COVID-19: a practice of building a community with a shared future for mankind. Infect Dis Poverty 2020 Apr 07;9(1):34 [FREE Full text] [doi: 10.1186/s40249-020-00650-1] [Medline: 32264957]
66. Dew K. Epidemics, panic and power: representations of measles and measles vaccines. Health (London) 2016 Jul 24;3(4):379-398. [doi: 10.1177/136345939900300403]
67. Petersen A, Lupton D. The New Public Health: Health and Self in the Age of Risk. Crows Nest, Australia: Allen & Unwin; 1996.
68. Alpher R. Comparing the corona to the Holocaust would not have occurred to any Israeli. Except for Netanyahu. Haaretz. 2020 Apr 22. URL: https://www.haaretz.co.il/gallery/television/tv-review/.premium-1.8788186 [accessed 2021-03-10]
69. PM Netanyahu's statement to the media on the Coronavirus. Israel Ministry of Foreign Affairs. 2020 Mar 25. URL: https://mfa.gov.il/MFA/PressRoom/2020/Pages/PM-Netanyahu-s-statement-to-the-media-on-the-Coronavirus-25-March-2020.aspx [accessed 2021-03-10]
70. van Asselt M, Vos E. Wrestling with uncertain risks: EU regulation of GMOs and the uncertainty paradox. J Risk Res 2008 Jan 1;11(1):281-300. [doi: 10.1080/13669870801990806]
71. van AM, Vos E, Rooijackers B. Science, knowledge and uncertainty in EU risk regulation. In: Inverson M, Vos E, editors. Uncertain Risks Regulated. Oxford, UK: Routledge-Cavendish; 2009:E-88.

72. Fox T. The role of uncertainty intolerance in European risk governance. 2009 Presented at: Risk Research Symposium; 2009; London, UK.
73. Gesser-Edelsburg A, Shir-Raz Y, Green MS. Why do parents who usually vaccinate their children hesitate or refuse? General good vs. individual risk. J Risk Res 2014 Dec 12;19(4):405-424. [doi: 10.1080/13669877.2014.983947]
74. Griffin RJ, Dunwoody S, Neuwirth K. Proposed model of the relationship of risk information seeking and processing to the development of preventive behaviors. Environ Res 1999 Feb;80(2 Pt 2):S230-S245. [doi: 10.1006/enrs.1998.3940] [Medline: 10092438]
75. Huurne ET, Gutteling J. Information needs and risk perception as predictors of risk information seeking. J Risk Res 2008 Oct;11(7):847-862. [doi: 10.1080/13669870701875750]
76. Kahlor L. PRISM: a planned risk information seeking model. Health Commun 2010 Jun;25(4):345-356. [doi: 10.1080/10410231003775172] [Medline: 20512716]
77. Witte K. Putting the fear back into fear appeals: The extended parallel process model. Commun Monogr 1992 Dec;59(4):329-349. [doi: 10.1080/03637759209376276]
78. De Vocht M, Claeys A, Cauberghe V, Uyttendaele M, Sas B. Won't we scare them? The impact of communicating uncontrollable risks on the public's perception. J Risk Res 2014 Nov 28;19(3):316-330. [doi: 10.1080/13669877.2014.971336]
79. Lofstedt RE. How can we make food risk communication better: where are we and where are we going? J Risk Res 2006 Dec;9(8):869-890. [doi: 10.1080/13669870601065585]
80. Palenchar MJ, Heath RL. Another part of the risk communication model: analysis of communication processes and message content. J Public Relat Res 2002 Mar;14(2):127-158. [doi: 10.1207/s1532754xjprr1402_3]
81. Slovic P. Beyond numbers: a broader perspective on risk perceptionrisk communication. In: Mayo DG, Hollander RD, editors. Acceptable Evidence: Science and Values in Risk Management. New York, NY: Oxford University Press; 1991:48-65.
82. Sandman PM. Ebola in the U.S. (so far): the public health establishment and the quarantine debate. The Peter Sandman Risk Communication Website. 2014 Nov 15. URL: http://www.psandman.com/col/Ebola-4.htm [accessed 2021-03-10]
83. Gonsalves G, Staley P. Panic, paranoia, and public health--the AIDS epidemic's lessons for Ebola. N Engl J Med 2014 Dec 18;371(25):2348-2349. [doi: 10.1056/NEJMp1413425] [Medline: 25372947]
84. Viebeck E. Health officials struggle to control the media narrative about Ebola. The Hill. 2014 Oct 12. URL: https://thehill.com/policy/healthcare/220458-feds-struggle-to-control-the-media-narrative-about-ebola [accessed 2021-03-10]
85. McKenna M. Ebolanoia: the only thing we have to fear is Ebola itself. Wired. 2014 Oct 22. URL: https://www.wired.com/2014/10/ebolanoia/ [accessed 2021-03-10]
86. De Vocht M, Claeys A, Cauberghe V, Uyttendaele M, Sas B. Won't we scare them? The impact of communicating uncontrollable risks on the public's perception. J Risk Res 2014 Nov 28;19(3):316-330. [doi: 10.1080/13669877.2014.971336]
87. Drabek TE, Erikson K. A new species of trouble: explorations in disaster, trauma, and community. Contemp Sociol 1994 Jul;23(4):507. [doi: 10.2307/2076356]
88. Witte K. Fear control and danger control: a test of the extended parallel process model (EPPM). Commun Monogr 2009 Jun 02;61(2):113-134. [doi: 10.1080/03637759409376328]
89. Bennett P, Calman K, Curtis S, Fischbacher-Smith D. Risk Communication and Public Health. Oxford, UK: Oxford University Press; 2010.
90. Thompson T, Parrott R, Nussbaum J. The Routledge Handbook of Health Communication, Second Edition. New York, NY: Routledge; 2011.
91. Kreuter MW, Strecher VJ, Glassman B. One size does not fit all: the case for tailoring print materials. Ann Behav Med 1999;21(4):276-283. [doi: 10.1007/BF02895958] [Medline: 10721433]
92. Lee N, Kotler P. Social Marketing Influencing Behaviors for Good, Fourth Edition. Thousand Oaks, CA: Sage Publications; 2011.
93. Serra YA, Colón-López V, Savas LS, Vernon SW, Fernández-Espada N, Vélez C, et al. Using intervention mapping to develop health education components to increase colorectal cancer screening in Puerto Rico. Front Public Health 2017;5:324 [FREE Full text] [doi: 10.3389/fpubh.2017.00324] [Medline: 29270400]
94. Fernandez ME, Ruiter RAC, Markham CM, Kok G. Intervention mapping: theory- and evidence-based health promotion program planning: perspective and examples. Front Public Health 2019;7:209 [FREE Full text] [doi: 10.3389/fpubh.2019.00209] [Medline: 31475126]
95. Schmid KL, Rivers SE, Latimer AE, Salovey P. Targeting or tailoring? Mark Health Serv 2008;28(1):32-37 [FREE Full text] [Medline: 18389854]
96. Lustria M. Message tailoring in health and risk messaging. In: Oxford Encyclopedia of Communication. Oxford, UK: Oxford University Press; 2017.
97. Kotler P, Lee N. Social Marketing: Influencing Behaviors for Good, 3rd Edition. Thousand Oaks, CA: Sage Publications; 2008.
98. Kreps G, Kunimoto E. Effective Communication in Multicultural Health Care Settings. Newbury Park, CA: Sage; 1994.
99. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]

100. Campbell A. Jacinda Ardern's coronavirus plan is working because, unlike others, she's behaving like a true leader. Independent. 2020 Apr 11. URL: https://www.independent.co.uk/voices/coronavirus-new-zealand-jacinda-ardern-cases-deaths-leadership-a9460591.html [accessed 2021-03-10]
101. Guttman N, Salmon CT. Guilt, fear, stigma and knowledge gaps: ethical issues in public health communication interventions. Bioethics 2004 Nov;18(6):531-552. [doi: 10.1111/j.1467-8519.2004.00415.x] [Medline: 15580723]
102. Frewer L, Hunt S, Brennan M, Kuznesof S, Ness M, Ritson C. The views of scientific experts on how the public conceptualize uncertainty. J Risk Res 2003 Jan;6(1):75-85. [doi: 10.1080/1366987032000047815]
103. Christley RM, Mort M, Wynne B, Wastling JM, Heathwaite AL, Pickup R, et al. "Wrong, but useful": negotiating uncertainty in infectious disease modelling. PLoS One 2013 Oct 16;8(10):e76277 [FREE Full text] [doi: 10.1371/journal.pone.0076277] [Medline: 24146851]
104. Slovic P. The construction of preference. Am Psychol 1995 May;50(5):364-371. [doi: 10.1037/0003-066x.50.5.364]
105. Palenchar MJ, Heath RL. Another part of the risk communication model: analysis of communication processes and message content. J Public Relat Res 2002 Mar;14(2):127-158. [doi: 10.1207/s1532754xjprr1402_3]
106. Sandman PM, Lanard J. Explaining and proclaiming uncertainty: risk communication lessons from Germany's deadly E. coli outbreak. The Peter Sandman Risk Communication Website. 2011 Aug 14. URL: http://www.psandman.com/col/GermanEcoli.htm [accessed 2021-03-10]
107. Gesser-Edelsburg A, Shir-Raz Y. Science vs. fear: the Ebola quarantine debate as a case study that reveals how the public perceives risk. J Risk Res 2015 Nov 05;20(5):1-23. [doi: 10.1080/13669877.2015.1100659]
108. Gesser-Edelsburg A, Shir-Raz Y, Green MS. Why do parents who usually vaccinate their children hesitate or refuse? General good vs. individual risk. J Risk Res 2014 Dec 12;19(4):405-424. [doi: 10.1080/13669877.2014.983947]
109. Gesser-Edelsburg A, Shir-Raz Y, Bar-Lev OS, James JJ, Green MS. Outbreak or epidemic? How Obama's language choice transformed the Ebola outbreak into an epidemic. Disaster Med Public Health Prep 2016 Aug;10(4):669-673. [doi: 10.1017/dmp.2016.48] [Medline: 27098568]
110. Heath R, O'Hair H. Handbook of Risk and Crisis Communication. New York, NY: Routledge; 2009.
111. Gesser-Edelsburg A, Mordini E, James JJ, Greco D, Green MS. Risk communication recommendations and implementation during emerging infectious diseases: a case study of the 2009 H1N1 Influenza Pandemic. Disaster Med Public Health Prep 2014 Apr 15;8(2):1-12. [doi: 10.1017/dmp.2014.27] [Medline: 24735822]
112. Brennan B, Hall W. Putting planning into practice: the communications response to H1N1. 2009 Presented at: A Global Communications Conference sponsored by the Pan American Health Organization and the United States Department of Health and Human Services; July 22, 2009; Washington, DC URL: https://www1.paho.org/cdmedia/riskcommguide/Putting%20Planning%20into%20Practice_GlobalComm_FinalReport_Eng.pdf
113. Fischhoff B. Scientifically sound pandemic risk communication (House Science Committee Briefing, Gaps in the National Flu Preparedness Plan: Social Science Planning and Response, December). In: New KD, editor. McGraw Hill Handbook of Terrorism and Counter-terrorism. New York, NY: McGraw Hill; 2005.
114. Driedger SM, Cooper E, Jardine C, Furgal C, Bartlett J. Communicating risk to aboriginal peoples: first nations and Metis responses to H1N1 risk messages. PLoS One 2013;8(8):e71106 [FREE Full text] [doi: 10.1371/journal.pone.0071106] [Medline: 23940697]
115. Gray L, MacDonald C, Mackie B, Paton D, Johnston D, Baker MG. Community responses to communication campaigns for influenza A (H1N1): a focus group study. BMC Public Health 2012 Mar 19;12(1):205 [FREE Full text] [doi: 10.1186/1471-2458-12-205] [Medline: 22429559]
116. Massey PD, Miller A, Saggers S, Durrheim DN, Speare R, Taylor K, et al. Australian Aboriginal and Torres Strait Islander communities and the development of pandemic influenza containment strategies: community voices and community control. Health Policy 2011 Dec;103(2-3):184-190. [doi: 10.1016/j.healthpol.2011.07.004] [Medline: 21868121]
117. Massey PD, Pearce G, Taylor KA, Orcher L, Saggers S, Durrheim DN. Reducing the risk of pandemic influenza in Aboriginal communities. Rural Remote Health 2009;9(3):1290 [FREE Full text] [Medline: 19728766]
118. Rudge S, Massey PD. Responding to pandemic (H1N1) 2009 influenza in Aboriginal communities in NSW through collaboration between NSW Health and the Aboriginal community-controlled health sector. N S W Public Health Bull 2010;21(1-2):26-29 [FREE Full text] [doi: 10.1071/NB09040] [Medline: 20374691]
119. Lai L, To W. Content analysis of social media: a grounded theory approach. J Electron Commer Res 2015;16(2):138-152 [FREE Full text]
120. Airoldi M. Ethnography and the digital fields of social media. Int J Soc Res Methodol 2018 Apr 25;21(6):661-673. [doi: 10.1080/13645579.2018.1465622]
121. Baxendale S, O'Toole A. Epilepsy myths: alive and foaming in the 21st century. Epilepsy Behav 2007 Sep;11(2):192-196. [doi: 10.1016/j.yebeh.2007.04.019] [Medline: 17602883]
122. Manyibe E. Top ten cancer myths. J Appl Rehabil Couns 2006;37(2):36.
123. Miller K. Book review: Girls' Violence: Myths and Realities. Youth Violence Juv Justice 2016 Aug 13;5(2):207-210. [doi: 10.1177/1541204006296166]
124. Nyhan B, Reifler J. Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. Vaccine 2015 Jan 09;33(3):459-464. [doi: 10.1016/j.vaccine.2014.11.017] [Medline: 25499651]

125. Nyhan B, Reifler J, Richey S, Freed GL. Effective messages in vaccine promotion: a randomized trial. Pediatrics 2014 Apr 03;133(4):e835-e842. [doi: 10.1542/peds.2013-2365] [Medline: 24590751]
126. Tasnim S, Hossain MM, Mazumder H. Impact of rumors and misinformation on COVID-19 in social media. J Prev Med Public Health 2020 May;53(3):171-174 [FREE Full text] [doi: 10.3961/jpmph.20.094] [Medline: 32498140]
127. Ali S. Hu Arenas 2020 Oct 07. [doi: 10.1007/s42087-020-00139-1]
128. Singhal A. The positive deviance approach to designing and implementing health communication interventions. In: Kim DK, Singhal A, Kreps GL, editors. Health Communication Strategies for Developing Global Health Programs. New York, NY: Peter Lang; 2013:176.
129. Singhal A, Greiner K. Using the positive deviance approach to reduce hospital-acquired infections at the veterans administration healthcare system in Pittsburgh. In: Suchman AL, Sluyter DG, Williamson PR, editors. Leading Change in Healthcare: Transforming Organizations Using Complexity, Positive Psychology, and Relationship-Centered Care. London, UK: Radcliffe Publishing; 2011:177.
130. MRSA Field Guide. Archived at the Internet Archive. Maryland Patient Safety Center. 2007. URL: https://web.archive.org/web/20160414013932/http://www.positivedeviance.org/pdf/manuals/Field%20Guide%20Sections%201-7.pdf [accessed 2021-03-10]
131. Gesser-Edelsburg A, Cohen R, Halavi AM, Zemach M, van Heerden PV, Sviri S, et al. Beyond the hospital infection control guidelines: a qualitative study using positive deviance to characterize gray areas and to achieve efficacy and clarity in the prevention of healthcare-associated infections. Antimicrob Resist Infect Control 2018;7:124 [FREE Full text] [doi: 10.1186/s13756-018-0418-x] [Medline: 30386593]
132. Lindberg C, Buscell P, Singhal A. Inviting Everyone: Healing Healthcare through Positive Deviance. Bordentown, NJ: Plexus Press; 2010.
133. Singhal A, Buscell P, Lindberg C. Inspiring Change and Saving Lives: The Positive Deviance Way. Bordentown, NJ: Plexus Press; 2014.
134. Lee H, Fawcett J, DeMarco R. Storytelling/narrative theory to address health communication with minority populations. Appl Nurs Res 2016 May;30:58-60. [doi: 10.1016/j.apnr.2015.09.004] [Medline: 27091254]
135. Larson HJ. The biggest pandemic risk? Viral misinformation. Nature 2018 Oct 16;562(7727):309-309. [doi: 10.1038/d41586-018-07034-4] [Medline: 30327527]
136. The power of proof: an evaluation primer. Tobacco Technical Assistance Consortium. 2020 Jul 23. URL: http://tacenters.emory.edu/resources/products_tools/tobacco/power-of-proof/ [accessed 2021-03-10]
137. Andreasen A. Marketing Social Change: Changing Behavior to Promote Health, Social Development, and the Environment. San Francisco, CA: Jossey-Bass; 1995.

Abbreviations

EPPM: extended parallel process model
PD: positive deviance

Edited by G Eysenbach, C Basch; submitted 11.10.20; peer-reviewed by A Chang, Q Wang; comments to author 05.11.20; revised version received 09.11.20; accepted 05.03.21; published 15.03.21.

Please cite as:
Gesser-Edelsburg A
Using Narrative Evidence to Convey Health Information on Social Media: The Case of COVID-19
J Med Internet Res 2021;23(3):e24948
URL: https://www.jmir.org/2021/3/e24948
doi:10.2196/24948
PMID:33674257

Viewpoint

Googling for Neurological Disorders: From Seeking Health-Related Information to Patient Empowerment, Advocacy, and Open, Public Self-Disclosure in the Neurology 2.0 Era

Mariano Martini[1], PhD; Nicola Luigi Bragazzi[1,2], MPH, MD, PhD

[1]Department of Health Sciences, School of Public Health, University of Genoa, Genoa, Italy

[2]Laboratory for Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto, ON, Canada

Corresponding Author:
Nicola Luigi Bragazzi, MPH, MD, PhD
Department of Health Sciences
School of Public Health
University of Genoa
Via Antonio Pastore 1
Genoa, 16132
Italy
Phone: 39 3279769965
Email: robertobragazzi@gmail.com

Abstract

Since its introduction, the internet has played a major role in reshaping patient-physician communication and interactions, having fostered a shift from a paternalistic to a patient-centered model. Because of its dynamic nature, the internet has been used as a platform to not only disseminate knowledge—favored by improved access to an increasing wealth of available resources—but also to spread advocacy and awareness, contribute to fund-raising, and facilitate open, public self-disclosure of one's own disease, thus eliminating any taboo and reducing the stigma associated with it. The era of Medicine 2.0 is characterized by openness, collaboration, participation, and social networking. The current situation is completely different from the time when Lorenzo Odone's parents, after his diagnosis of adrenoleukodystrophy, decided to attend medical school in order to collect information about a devastating, unknown disease and had to contend with medical authorities at that establishment to convince them of the alleged effectiveness and safety of their discovered therapeutics. Orphan and rare neurological diseases have currently received recognition on web-based resources. However, while the intention is not to ridicule Odone's family legacy and the "complicated lessons" they have reported, some issues should be carefully addressed by health authorities, such as the reputability, reliability, and accuracy of material available on the internet and prevention of the dissemination of material that could instill illusions and unjustified hopes in individuals seeking medical treatment. Neurologists should be aware of such digital resources, participate in web-based activities, and recommend select high-quality websites to their patients.

(J Med Internet Res 2021;23(3):e13999) doi:10.2196/13999

KEYWORDS

advocacy; health information seeking; neurological disorders; open self-disclosure

Introduction

In 1984, Lorenzo Odone was diagnosed with adrenoleukodystrophy, a severe neurodegenerative disorder. His parents, Michaela and Augusto Odone, decided not to resign to unbearable and devastating pain but rather to study medicine, and—after spending much time in the libraries of the National Institutes of Health—they devised a putative treatment strategy (the so-called "Lorenzo's oil"). Since then, 30 years have elapsed, and the situation has changed. Markedly more information is available on adrenoleukodystrophy and other rare neurological diseases; the number of voluntary health associations has notably increased, together with patient awareness; and owing to the growing availability of web-based medical material, the material that was relegated and confirmed within the university precinct and written in an obscure, technical language at the time of Odone's family is publicly available.

Although Odone's heritage has been undoubtedly immense, in terms of both pharmaceutical legacy and the challenge they posed to the traditional physician-patient relationship [1], ethical

issues concerning the clinical effectiveness and the safety profile of new experimental therapies should be carefully considered.

Improved access to health care information has reshaped the concept of medical paternalism and has caused a shift from an "informed patient care" to a "patient-informed care" [2].

The internet and the new communication and information technologies (ICTs) have enabled physicians to deliver care remotely (the so-called "teleneurology") [1] and have provided new tools and opportunities for disseminating and improving resident education [3]. However, undoubtedly, the major novelty lies in the profound changes associated with medical communication in the Copernican revolution: indeed, ICTs represent one of the elements of the new conceptual framework of P6 medicine, where the 6 Ps stand for "personalized," "predictive," "preventive," "participatory," "psycho-cognitive," and "public" [4-6].

Managing the Complexity of Neurological Disorders: Online Health Communities

Neurological diseases are particularly complex, multifactorial pathologies that require disease-specific expertise, which is rarely found in a single specialist. For neurological diseases, comprehensive management including different professionals—ranging from a neurologist to a physical therapist, occupational therapist, and speech therapist—is recommended. In this context, the concept of online health communities introduced by van der Heijk et al [7] is of particular value, specifically in the treatment of chronic diseases. ICTs can facilitate highly integrated, shared, multidisciplinary management and high-quality, affordable physician-patient interaction.

For instance, ParkinsonNet [8] represents a unique innovation developed as a network of physicians in several regions in the Netherlands, and it has numerous proven benefits and positive outcomes, including cost-effectiveness, increase in health literacy and disease knowledge, and compliance to treatment.

Patients' Willingness and Acceptance of the Internet and New ICTs

Patients with neurological disorders appear to actively use the internet to search for information on their pathology. They (or their relatives and parents) [9] look for resources on their prognoses, outcomes, and treatment [10] or surf the web in order find an expert specialist available for consultation [11].

One reason why internet usage is particularly widespread among patients could be linked to the fact that, because of the aging population and the increasing number of diseases and patients, physicians dedicate lesser time to them patients than before, and patients feel that their doubts and questions are not sufficiently addressed. For example, Hoch et al [12] reported that up to 20% of the users of the epilepsy "Webforum" claimed they did not receive adequate information from their physicians.

Chiò et al [13] conducted a survey among patients with amyotrophic lateral sclerosis and their caregivers and reported that approximately 55% and 83.3% of them, respectively, surfed the web for seeking health-related information. However, they were rather critical of the quality of the web-based material.

According to a survey carried out by Haase et al [14], most patients with multiple sclerosis own a personal computer and use it quite regularly. However only 20% of them would use mobile phones to communicate with their physicians, 40% would use the internet, 54% would use email, and approximately 68% would use at least one type of electronic communication device. Since the use of electronic tools was found to be a significant predictor of the acceptance of electronic interactions with health care providers, studies should particularly focus on increasing patients' usage of new technologies and understanding which factors may act as barriers and attempting to eliminate them.

Few studies have focused on these barrier parameters. Nielsen et al [15] reported that ethnicity, vision impairment, and arm and hand disabilities markedly inhibit the use of web-based technologies in a cohort of patients with multiple sclerosis. They suggest that technological adaptation (such as voice-driven commands and text written using enlarged fonts) should be ensured in order to increase internet usage among patients.

Another potential barrier is cultural and is associated with the use of technical language, being usually adopted by medical resources. Elliott and Shneker [16] reported that only 3% of the Epilepsy Foundation's website [17], a portal entirely devoted to seizures, adhered to the recommendations of the Institute of Medicine and the US Department of Education that health-related information should be simple and clearly written in order to be understood by any user.

Sharing One's Own Experience and Reducing the Stigma: the Phenomenon of Open, Public Self-Disclosure

Ad hoc websites designed specifically for patients, such as PatientsLikeMe [18] or The Italian headache disorders website [19], help patients share their experiences and guide them through their difficulties in decision-making. Patients are not mere passive subjects but rather active producers and consumers at the same time (the prosumer model). Iaconesi [20] has exploited the potential of new technologies (open-source software, file uploading and sharing, and commenting and posting) to publicly self-disclose his brain cancer, thus becoming the emblematic hallmark of the new P6 medicine, reducing the associated stigma, and providing novel insights into the treatment and management of chronic neurodegenerative disorders.

The Issue of Reliability and the Quality of the Content of Web-Based Material

An important issue associated with web-based material on neurological diseases is its quality, reputability, and reliability. Information should be written in a clear manner, providing accuracy, referenced details, and updated content. The websites should be cured in a manner that ownership, authorship,

sponsorship, funding and financial support, and any other potential conflicts of interest are properly disclosed. It would be ideal for websites to be tailored to meet individual patients' requirements, providing, for example, the opportunity to interact with experts and clinicians [21].

Peterlin et al [21] indicated high-quality and excellent websites devoted to cluster headaches, which can be important resources for patients, even though the overall quality of websites dedicated to headaches is generally poor and mediocre [22]. In total, 72.5% of these headache-related websites contained advertisements and most of them contained technical information. Efforts should be made to promote the more reputed websites, which may sometimes be difficult to come across by the general public.

Hoch et al [12] reported that only 6% of the information available on a website dedicated to epilepsy was inaccurate, whereas Di Pietro et al [23] reported that websites on neurodevelopmental disorders contained misleading information and, at least in 20% of cases, quoted scientific references in an incorrect or irrelevant manner.

Moreover, some of the online material could not be based on sufficient scientific evidence, as in the case of multiple sclerosis, for which some websites have described chronic cerebrospinal venous insufficiency (CCSVI) as the main pathogenic factor without clearly stating that this is still controversial and no scientific consensus has been reached. Instead, patients mining these websites should be advised of the fact that no controlled randomized clinical trials have been performed in order to confirm and replicate this finding. Fragoso [24] has maintained that this fraudulent ideology of describing the CCSVI theory as the "liberation treatment" could instill unjustified hopes and illusions in patients. Indeed, Bragazzi [25] reported that "CCSVI" is a highly searched term, the volume trend of which correlated with the volume trend of searching "multiple sclerosis" as a keyword.

Pucci [26] reported that the 32% of the web-based material on multiple sclerosis and Alzheimer disease was unreliable and misleading, and patients requested an intervention by a physician. Hence, it is important to conduct content analysis of web-based material. Clinicians should remain abreast of such content analyses in order to be prepared to discuss them with their patients. Furthermore, Di Pietro et al [23] suggested that "new partnerships between advocacy and experts" may ensure quality material and avoid spreading disinformation among patients.

Health authorities and organizations should establish clear standards to be followed and should monitor their adherence; this should be the onus of the medical establishment and should not be overlooked. Professionals themselves can increase their presence on web-based platforms, being active authors of weblogs or, at least, participating in the generation of websites [27].

Positive and Negative Aspects of the New Technologies

Currently, the internet has a dual nature. It is important to emphasize that, if, on the one hand, the ICTs can positively contribute to the treatment and management of neurological disorders, on the other hand, they can have a negative impact. Regarding the positive aspects, ICTs can favor rapid, real-time dissemination of information; disseminate an unprecedented wealth of web-based relevant material; reduce stigma and discrimination; enhance social and peer support; develop new forms of physician-patient interaction and communication, which are truly patient-centered and personalized; increase health literacy; improve self-awareness and self-empowerment among patients; spread advocacy; and contribute to fund-raising. However, ICTs can also divulge potentially misleading and dangerous information regarding the etiology and management of neurological disorders, proposing ineffective and unsafe treatments. Web-based material can, indeed, be of poor quality, inaccurate, and not based on scientific evidence. Unfortunately, in the era of "fake news" and in the "post-truth age," the risk of spreading unreliable information is quite valid [28].

Conclusions

The internet has raised patient's awareness, facilitating the initiation and spread of self-help movements in the so called "electronic peer-to-peer virtual communities" [19,29]. It has profoundly changed the patient-physician relationship [26]; therefore, it is critical for clinicians to be fully aware of these phenomena and attempt to exploit them. For example, physicians could directly participate in developing dedicated websites or could at least have discussions with their patients about their internet usage and activities, encouraging them to properly surf the internet [26] and recommending to them a list of select high-quality websites [21,22].

Patients themselves are sometimes aware of the poor nature of the web-based material, concurrent with the findings of Marrie et al [30] that approximately 40% of patients with multiple sclerosis had concerns regarding the quality of information they obtained on the internet.

Monitoring of keywords and search hits by patients [31-39] could help clinicians understand the patients' requirements. Predictors of internet usage and digital activities are generally associated with age, degree of symptom severity and neurological impairment, and socioeconomic status [29].

Finally, a proper understanding, the elimination of barriers to accessing information on web-based platforms, and the regulation of information—ensuring high quality standards—are important for facilitating internet usage in the era of Medicine 2.0 [40].

Conflicts of Interest

None declared.

References

1. Larner AJ. Teleneurology: an overview of current status. Pract Neurol 2011 Oct;11(5):283-288. [doi: 10.1136/practneurol-2011-000090] [Medline: 21921003]
2. Gardiner R. The transition from 'informed patient' care to 'patient informed' care. Stud Health Technol Inform 2008;137:241-256. [Medline: 18560085]
3. Fung K, Tihan T. Internet and World Wide Web-based tools for neuropathology practice and education. Brain Pathol 2009 Apr;19(2):323-331. [doi: 10.1111/j.1750-3639.2009.00263.x] [Medline: 19290999]
4. Bragazzi N. From P0 to P6 medicine, a model of highly participatory, narrative, interactive, and "augmented" medicine: some considerations on Salvatore Iaconesi's clinical story. Patient Prefer Adherence 2013;7:353-359 [FREE Full text] [doi: 10.2147/PPA.S38578] [Medline: 23650443]
5. Bragazzi NL. Children, adolescents, and young adults participatory medicine: involving them in the health care process as a strategy for facing the infertility issue. Am J Bioeth 2013;13(3):43-44. [doi: 10.1080/15265161.2012.760674] [Medline: 23428039]
6. Bragazzi NL, Del Puente G. Why P6 Medicine Needs Clinical Psychology and a Trans-Cultural Approach. Health Psychol Res 2013 Jan 02;1(1):e5 [FREE Full text] [doi: 10.4081/hpr.2013.e5] [Medline: 26973894]
7. van der Eijk M, Faber MJ, Aarts JW, Kremer JA, Munneke M, Bloem BR. Using online health communities to deliver patient-centered care to people with chronic conditions. J Med Internet Res 2013 Jun 25;15(6):e115 [FREE Full text] [doi: 10.2196/jmir.2476] [Medline: 23803284]
8. ParkinsonNet. URL: https://www.parkinsonnet.com/ [accessed 2021-03-23]
9. Lu C, Wirrell E, Blackman M. Where do families of children with epilepsy obtain their information? J Child Neurol 2005 Nov;20(11):905-910. [doi: 10.1177/08830738050200110801] [Medline: 16417861]
10. Henson JW. Neurology patients online: perceptions and reality. Nat Clin Pract Neurol 2008 Jul;4(7):347. [doi: 10.1038/ncpneuro0852] [Medline: 18600254]
11. Imai N, Yagi N, Konishi T, Serizawa M, Kobari M. Websites offer helpful information concerning consultation with headache specialists. Cephalalgia 2010 Apr;30(4):496-499. [doi: 10.1111/j.1468-2982.2009.01915.x] [Medline: 19515123]
12. Hoch D, Norris D, Lester J, Marcus A. Information exchange in an epilepsy forum on the World Wide Web. Seizure 1999 Feb;8(1):30-34 [FREE Full text] [doi: 10.1053/seiz.1998.0217] [Medline: 10091845]
13. Chiò A, Montuschi A, Cammarosano S, De Mercanti S, Cavallo E, Ilardi A, et al. ALS patients and caregivers communication preferences and information seeking behaviour. Eur J Neurol 2008 Jan;15(1):55-60. [doi: 10.1111/j.1468-1331.2007.02000.x] [Medline: 18005051]
14. Haase R, Schultheiss T, Kempcke R, Thomas K, Ziemssen T. Use and acceptance of electronic communication by patients with multiple sclerosis: a multicenter questionnaire study. J Med Internet Res 2012 Oct 15;14(5):e135 [FREE Full text] [doi: 10.2196/jmir.2133] [Medline: 23069209]
15. Nielsen AS, Halamka JD, Kinkel RP. Internet portal use in an academic multiple sclerosis center. J Am Med Inform Assoc 2012;19(1):128-133 [FREE Full text] [doi: 10.1136/amiajnl-2011-000177] [Medline: 21571744]
16. Elliott JO, Shneker BF. A health literacy assessment of the epilepsy.com website. Seizure 2009 Jul;18(6):434-439 [FREE Full text] [doi: 10.1016/j.seizure.2009.04.003] [Medline: 19419887]
17. epilepsy.com. The Epilepsy Foundation. URL: epilepsy.com [accessed 2021-03-23]
18. PatientsLikeMe. URL: https://www.patientslikeme.com/ [accessed 2021-03-23]
19. Rossi P, Sances G, Nappi G. www.cefalea.it: the first five years. J Headache Pain 2005 Sep;6(4):264-267 [FREE Full text] [doi: 10.1007/s10194-005-0203-0] [Medline: 16362682]
20. Iaconesi S. La Cura, An Open Source Cure for Cancer. Big Data 2013 Sep;1(3):124-129. [doi: 10.1089/big.2013.0028] [Medline: 27442192]
21. Peterlin BL, Gambini-Suárez E, Levin M. Cluster headache: a review of online resources for patients and clinicians. Curr Pain Headache Rep 2007 Apr;11(2):148-153. [doi: 10.1007/s11916-007-0013-y] [Medline: 17367595]
22. Peterlin B, Gambini-Suarez E, Lidicker J, Levin M. An analysis of cluster headache information provided on internet websites. Headache 2008 Mar;48(3):378-384. [doi: 10.1111/j.1526-4610.2007.00951.x] [Medline: 18005143]
23. Di Pietro NC, Whiteley L, Mizgalewicz A, Illes J. Treatments for neurodevelopmental disorders: evidence, advocacy, and the Internet. J Autism Dev Disord 2013 Jan;43(1):122-133. [doi: 10.1007/s10803-012-1551-7] [Medline: 22592952]
24. Fragoso YD. The internet racing ahead of the scientific evidence: the case of "liberation treatment" for multiple sclerosis. Arq Neuropsiquiatr 2011 Jun;69(3):525-527 [FREE Full text] [doi: 10.1590/s0004-282x2011000400022] [Medline: 21755134]
25. Bragazzi NL. Infodemiology and infoveillance of multiple sclerosis in Italy. Mult Scler Int 2013;2013:924029 [FREE Full text] [doi: 10.1155/2013/924029] [Medline: 24027636]
26. Pucci E. Is the internet transforming the physician-consumer relationship? Preliminary data in a neurological setting. Eur J Neurol 2003 Mar;10(2):192. [doi: 10.1046/j.1468-1331.2003.00502_5.x] [Medline: 12603300]
27. Lagu T, Kaufman EJ, Asch DA, Armstrong K. Content of weblogs written by health professionals. J Gen Intern Med 2008 Oct;23(10):1642-1646 [FREE Full text] [doi: 10.1007/s11606-008-0726-6] [Medline: 18649110]

28. Lavorgna L, Lanzillo R, Brescia Morra V, Abbadessa G, Tedeschi G, Bonavita S. Social Media and Multiple Sclerosis in the Posttruth Age. Interact J Med Res 2017 Sep 27;6(2):e18 [FREE Full text] [doi: 10.2196/ijmr.7879] [Medline: 28954717]
29. Shavazi M, Morowatisharifabad M, Shavazi M, Mirzaei M, Ardekani A. Online Social Support for Patients with Multiple Sclerosis: A Thematic Analysis of Messages Posted to a Virtual Support Community. Int J Community Based Nurs Midwifery 2016 Jul;4(3):188-198 [FREE Full text] [Medline: 27382585]
30. Marrie RA, Salter AR, Tyry T, Fox RJ, Cutter GR. Preferred sources of health information in persons with multiple sclerosis: degree of trust and information sought. J Med Internet Res 2013 Mar 17;15(4):e67 [FREE Full text] [doi: 10.2196/jmir.2466] [Medline: 23635393]
31. Abedi V, Mbaye M, Tsivgoulis G, Male S, Goyal N, Alexandrov AV, et al. Internet-based information-seeking behavior for transient ischemic attack. Int J Stroke 2015 Dec;10(8):1212-1216. [doi: 10.1111/ijs.12593] [Medline: 26282685]
32. Bragazzi NL, Bacigaluppi S, Robba C, Nardone R, Trinka E, Brigo F. Infodemiology of status epilepticus: A systematic validation of the Google Trends-based search queries. Epilepsy Behav 2016 Feb;55:120-123. [doi: 10.1016/j.yebeh.2015.12.017] [Medline: 26773681]
33. Brigo F, Erro R. Why do people google movement disorders? An infodemiological study of information seeking behaviors. Neurol Sci 2016 May;37(5):781-787. [doi: 10.1007/s10072-016-2501-5] [Medline: 26846327]
34. Brigo F, Lochner P, Tezzon F, Nardone R. Web search behavior for multiple sclerosis: An infodemiological study. Mult Scler Relat Disord 2014 Jul;3(4):440-443. [doi: 10.1016/j.msard.2014.02.005] [Medline: 25877054]
35. Brigo F, Otte W, Igwe S, Ausserer H, Nardone R, Tezzon F, et al. Information-seeking behaviour for epilepsy: an infodemiological study of searches for Wikipedia articles. Epileptic Disord 2015 Dec;17(4):460-466 [FREE Full text] [doi: 10.1684/epd.2015.0772] [Medline: 26575365]
36. Brigo F, Trinka E. Google search behavior for status epilepticus. Epilepsy Behav 2015 Aug;49:146-149. [doi: 10.1016/j.yebeh.2015.02.029] [Medline: 25873438]
37. Lavi-Blau T, Ekstein D, Neufeld MY, Eyal S. Use of antiepileptic drugs during pregnancy and lactation: Type of information provided by searching Google. Epilepsy Behav 2016 Feb;55:113-119 [FREE Full text] [doi: 10.1016/j.yebeh.2015.12.002] [Medline: 26773680]
38. Milošević M, Friedrich L, Tomasović S, Bielen I. Online epilepsy counseling in Croatia: What do users want to know? Seizure 2016 Oct;41:116-119 [FREE Full text] [doi: 10.1016/j.seizure.2016.07.020] [Medline: 27525581]
39. Wang H, Chen D, Yu H, Chen Y. Forecasting the Incidence of Dementia and Dementia-Related Outpatient Visits With Google Trends: Evidence From Taiwan. J Med Internet Res 2015 Nov 19;17(11):e264 [FREE Full text] [doi: 10.2196/jmir.4516] [Medline: 26586281]
40. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008 Aug 25;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]

Abbreviations

CCSVI: chronic cerebrospinal venous insufficiency
ICT: communication and information technology

Edited by G Eysenbach; submitted 12.03.19; peer-reviewed by T Cruvinel, L Lavorgna; comments to author 03.04.19; revised version received 04.04.19; accepted 04.04.19; published 26.03.21.

Please cite as:
Martini M, Bragazzi NL
Googling for Neurological Disorders: From Seeking Health-Related Information to Patient Empowerment, Advocacy, and Open, Public Self-Disclosure in the Neurology 2.0 Era
J Med Internet Res 2021;23(3):e13999
URL: https://www.jmir.org/2021/3/e13999
doi:10.2196/13999
PMID:30946019

Original Paper

A Novel Mobile App (Heali) for Disease Treatment in Participants With Irritable Bowel Syndrome: Randomized Controlled Pilot Trial

Aaron J Rafferty[1], MSc; Rick Hall[2], PhD; Carol S Johnston[1], PhD

[1]College of Health Solutions, Arizona State University, Phoenix, AZ, United States

[2]Edson College of Nursing and Health Innovation, Arizona State University, Phoenix, AZ, United States

Corresponding Author:
Carol S Johnston, PhD
College of Health Solutions
Arizona State University
HLTHN 532 Phoenix Downtown Campus
Phoenix, AZ, 85004
United States
Phone: 1 602 496 2539
Email: Carol.johnston@asu.edu

Abstract

Background: A diet high in fermentable, oligo-, di-, monosaccharides and polyols (FODMAPs) has been shown to exacerbate symptoms of irritable bowel syndrome (IBS). Previous literature reports significant improvement in IBS symptoms with initiation of a low FODMAP diet (LFD) and monitored reintroduction. However, dietary adherence to the LFD is difficult, with patients stating that the information given by health care providers is often generalized and nonspecific, requiring them to search for supplementary information to fit their needs.

Objective: The aim of our study was to determine whether Heali, a novel artificial intelligence dietary mobile app can improve adherence to the LFD, IBS symptom severity, and quality of life outcomes in adults with IBS or IBS-like symptoms over a 4-week period.

Methods: Participants were randomized into 2 groups: the control group (CON), in which participants received educational materials, and the experimental group (APP), in which participants received access to the mobile app and educational materials. Over the course of this unblinded online trial, all participants completed a battery of 5 questionnaires at baseline and at the end of the trial to document IBS symptoms, quality of life, LFD knowledge, and LFD adherence.

Results: We enrolled 58 participants in the study (29 in each group), and 25 participants completed the study in its entirety (11 and 14 for the CON and APP groups, respectively). Final, per-protocol analyses showed greater improvement in quality of life score for the APP group compared to the CON group (31.1 and 11.8, respectively; *P*=.04). Reduction in total IBS symptom severity score was 24% greater for the APP group versus the CON group. Although this did not achieve significance (–170 vs –138 respectively; *P*=.37), the reduction in the subscore for bowel habit dissatisfaction was 2-fold greater for the APP group than for the CON group (*P*=.05).

Conclusions: This initial study provides preliminary evidence that Heali may provide therapeutic benefit to its users, specifically improvements in quality of life and bowel habits. Although this study was underpowered, findings from this study warrant further research in a larger sample of participants to test the efficacy of Heali app use to improve outcomes for patients with IBS.

Trial Registration: ClinicalTrials.gov NCT04256551; https://clinicaltrials.gov/ct2/show/NCT04256551

(J Med Internet Res 2021;23(3):e24134) doi:10.2196/24134

KEYWORDS

irritable bowel syndrome; artificial intelligence; mobile app; low FODMAP diet; randomized controlled trial

Introduction

Irritable bowel syndrome (IBS) is among the most prevalent functional gastrointestinal (GI) disorders and is characterized by a number of symptoms, including recurrent abdominal pain, altered bowel habits, bloating, and distention [1]. IBS affects roughly 15 million people within the United States [2,3] and results in yearly direct health care costs between US $30 and

$75 billion [4,5], while indirect costs are estimated at an additional US $20 billion a year. Individuals with IBS often experience psychological distress and anxiety from a lack of clear understanding of the condition, being told that symptoms are all in their heads, and feelings of not being heard, all of which further exacerbate symptoms and affect quality of life [1]. Additionally, affected individuals spend significant time seeking medical support and undergoing numerous assortments of tests that all lead to a loss of work, disability, lack of productivity, and increased mortality [4,6]. Furthermore, both central factors (eg, psychological, cognitive and neuro-hormonal) and peripheral factors (eg, gut flora, genetics, and diet) have been shown to exacerbate the severity of symptoms over time [7,8].

A diet high in fermentable, oligo-, di-, monosaccharides and polyols (FODMAPs) has been shown to exacerbate symptoms of IBS [9], and research suggests adherence to a low FODMAP diet (LFD) can improve IBS symptoms [10-19]. For this reason, an LFD has become a popular tool to manage IBS and IBS-like symptoms with comparable success rates to pharmacological methods [20]. However, adherence to an LFD is difficult as FODMAP compounds are present in a variety of fruits, vegetables, grains, dairy, meats, and condiments. Moreover, patients following an LFD state that information provided by medical practitioners is often generalized and nonspecific, requiring them to search for supplementary information to fit their individual needs [21]. Support from a multidisciplinary team has been shown to mitigate these barriers to treatment of functional bowel disorders [22]; however, accessibility becomes a concern when additional barriers arise for the patient, including the time and finances needed to accommodate additional support.

Mobile apps using artificial intelligence (AI) in consort with a multidisciplinary team within a platform are gaining traction as useful tools for supporting the management of chronic conditions like diabetes and hypertension [23-26]. However, an app designed to treat IBS symptoms using AI has yet to be explored. Of the apps developed for IBS treatment, the Constant-Care web app [27], developed by researchers in Denmark, has been used by participants to successfully monitor their IBS symptoms in dietary treatment studies [10-12,27,28]. However, this app was not offered as a mobile app and only provided symptom tracking and monitoring approaches without real-time assistance for patients to improve their LFD adherence. Instead, improvement in LFD adherence relied on in-person and online education modules that could not be used in real time to make decisions [27]. To date, an app using AI to reduce IBS symptoms and improve adherence to the LFD has not been tested.

Heali AI is a personalized nutrition software company, founded in 2018 in Los Angeles, California, by a team of software engineers and registered dietitians. The Heali app uses AI to scan menus and barcodes to provide nutrition information and recommendations in accordance with user-specific individualized diet plans. Typically, app users are matched to foods based on their selected dietary preferences using a traffic light system: green (foods that fit the diet plan), yellow (foods that can be consumed in moderation), and red (foods to avoid completely). The app can be programmed to focus on a wide range of dietary preferences, including micro- and macronutrient preferences, as well as specific preprogrammed evidence-based diets such as the LFD, vegan diet, gluten-free diet, or keto diet. Once programmed, the app then helps users find foods that align with their personalized dietary needs while eating out, at the grocery store, or at home in an effort to improve dietary adherence, symptom control, and quality of life.

The purpose of this per-protocol study was to determine whether Heali, a novel AI dietary app, reduces IBS symptoms through improving adherence to the LFD as compared to standard online dietary education in populations with IBS or IBS-like symptoms after 30 days of use. It was hypothesized that, compared to standard online education, the novel AI dietary app would improve the primary outcomes of IBS symptom severity and quality of life via improving the secondary outcomes, which include adherence to an LFD and dietary knowledge related to the LFD.

Methods

Participants

Participants were eligible for the study if they had moderate to severe IBS based on an IBS symptom severity scale (IBS-SSS) score of 175 or greater, met Rome IV criteria for IBS, and had IBS-like symptoms for the past 3 months or longer [29]. Diagnosis using the Rome IV criteria also classifies patients by symptomology: IBS with constipation (IBS-C), IBS with diarrhea (IBS-D), IBS mixed typed (IBS-M), or IBS unsubtyped (IBS-U).

Eligibility also required that participants be between 18 and 65 years of age and own a cell phone. Exclusion criteria included use of a dietary app or elimination diet (eg, LFD, specific carbohydrate, vegan, vegetarian, gluten free, dairy free) within the last 6 months, food allergies (not including food intolerances), smoking habit, history of chronic disease other than GI dysfunction (eg, diabetes, cardiovascular disease, hypertension, eating disorders, or diseases of the lungs, kidney, liver, or thyroid), or being a nutrition student or professional.

The study was advertised via online flyers, social media posts (Facebook, Twitter, Instagram), list servers, and recruitment services across a university community. The recruitment message indicated that participants must be willing to participate in an LFD intervention over a 30-day period with a 10-day pretrial monitoring period. The message also stated that participants would be instructed to complete online questionnaires on a biweekly basis from the start of the study. Participants provided consent via email, and the study was approved by the Arizona State University Institutional Review Board. The study was registered at ClinicalTrials.gov (NCT04256551).

A total of 58 participants were recruited. According to previous literature, 20 participants per group would provide 80% power to detect at least an 81-point decrease in IBS-SSS symptom scores between groups at an α value of .05 [10,16,18,30]. Thus, assuming a retention rate of 80%, a minimum of 25 participants per group was desired.

Study Design

This 40-day randomized controlled experimental study consisted of a 10-day baseline monitoring period followed by a 30-day intervention period. Prior to the start of the intervention, author AJR randomized participants into 2 groups in order of the date they enrolled into the study, using an online random number generator. Groups were defined as follows: the experimental group (APP), who had access to the AI dietary mobile app and standard dietary education materials, and the control group (CON), who only had access to standard dietary education materials. Researchers and participants were both unblinded to the intervention type. Both groups received standard online LFD intervention resources developed by the University of Michigan [31]. The online education provided a description of the LFD and instructions for adherence, a high and low FODMAP guide, and an LFD cooking guide that contained 21 sample meals. The APP group received the same educational materials in addition to access to the Heali mobile app. Standardized emails were sent out to the APP and CON groups, which provided resources and questionnaires to support and track participation. Upon completion of the study, participants were also provided a guide for the reintroduction of FODMAPs, the final stage of the low FODMAP diet, which they were able to tailor to their needs. Those who successfully completed the intervention were able to use the AI mobile application to support their reintroduction stage over a 6-month period after completion.

Heali Mobile App

For this study, APP participants only had access to the LFD, meaning they had access to all features within the app with the exception of the ability to further personalize their preferences beyond the LFD. They also had access to a within-app health coach, which provided participants weekly reminders to use the app and answered questions elicited by the participants. APP participants were instructed to use the app daily. Participants gained access to the app via a cell phone login and received a standard usage tutorial in PDF form.

Measures

Prior to the 10-day baseline period and during the screening process, participants completed 3 questionnaires, including a demographic screener (inclusion criteria and cofounding variables, such as physical activity [32], smoking, and history of disease), the IBS-SSS diagnostic tool [33], and the Rome IV screener [29]. All questionnaires were sent via Google Forms and deployed in an email to participants. Once accepted into the study, all participants were required to complete 5 questionnaires at the start of the 30-day intervention period and at the end of the trial, specifically the Rome IV questionnaire, the IBS-SSS diagnosis tool, the low FODMAP dietary consumption (LFDA) questionnaire, the low FODMAP dietary knowledge (LFDK) questionnaire, and a quality of life questionnaire. The 6-item Rome IV questionnaire (1-2 minutes to complete) is the current standard diagnostic tool developed by the Rome Foundation used to determine severity of GI dysfunction and diagnose IBS [29]. The 5-item IBS-SSS questionnaire (1-2 minutes to complete) is a 500-point symptom severity screener validated by Francis et al [33]. Respondents were categorized into 1 of 4 categories based on their responses: <75, no symptoms; 75 to <175, mild IBS; 175 to <300, moderate IBS; and ≥300, severe IBS [33]. The IBS-SSS was also completed once every 10 days over the course of the trial. The 110-item LFDA questionnaire (based on the NHANES food frequency questionnaire) recorded the number of times daily that respondents ate certain FODMAP items [34]. The total scoring range was 0-385, with 0 indicating no FODMAPs eaten in the last month and 385 indicating every FODMAP was eaten 2-3 times per day or more [34]. The 12-item LFDK questionnaire quantified the respondents' knowledge of LFD. The total scoring range was 0-60, with 60 being the best possible knowledge score. The survey was modified from a 12-item validated tool by Krause et al [35] to assess FODMAP knowledge rather than general nutrition knowledge. These modifications included the following: "When I have questions on FODMAP foods, I know where I can find information on this issue," where "FODMAP foods" was substituted for "healthy nutrition" for the purposes of this study. The 13-item quality of life measure represented domains 1 and 2 of the World Health Organization (WHO) quality of life questionnaire, in which the total scoring range was 0-200, with 200 being the best possible quality of life score [36].

Statistical Analysis

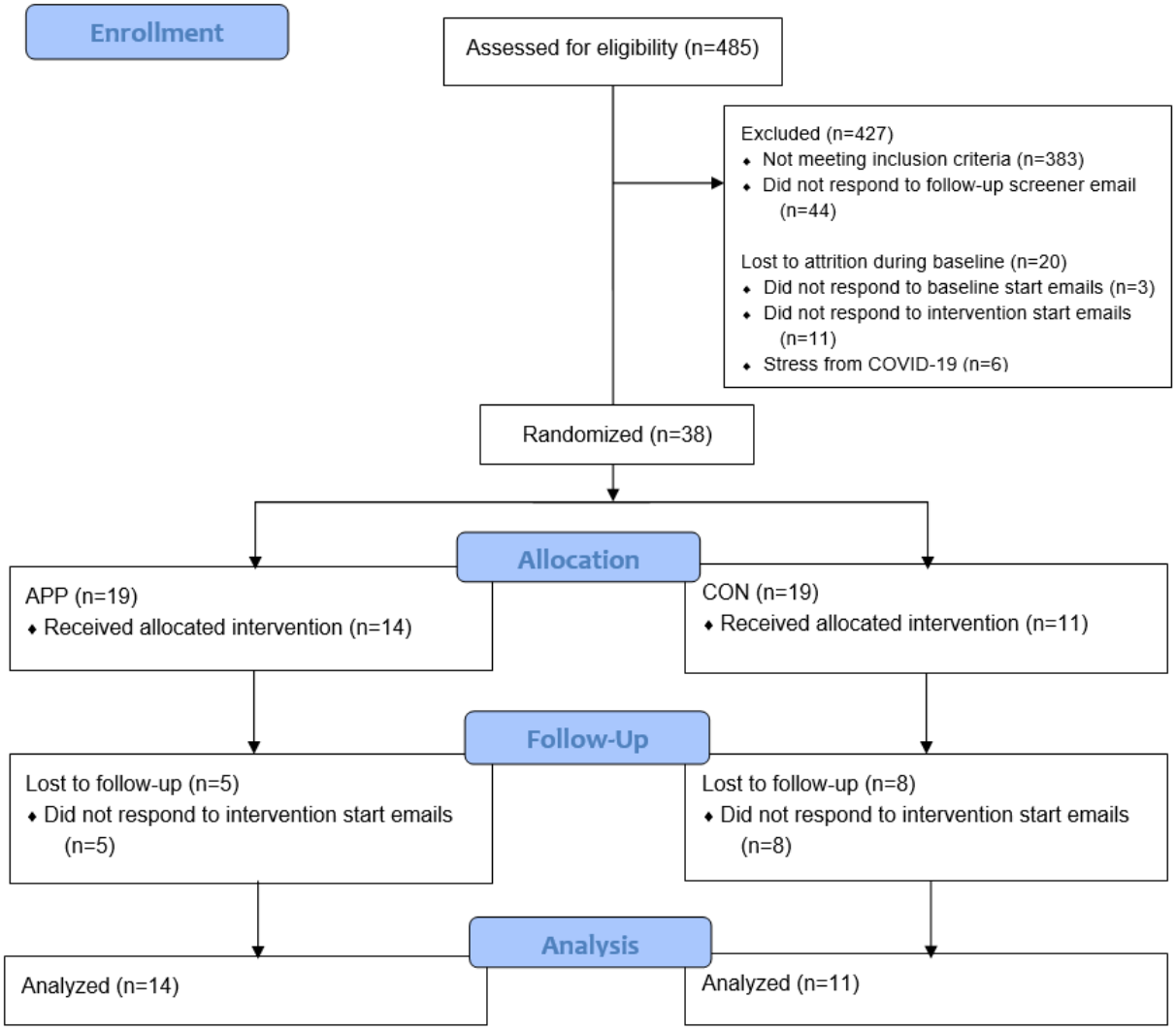
Data are presented as mean (SD). Analyses were conducted using SPSS Statistics Version 26 (IBM Corp) for Windows (Microsoft Corp). Spearman test was used to determine correlations in baseline variables. Mann-Whitney U tests were used to determine differences between groups. Pearson chi-square test was used to determine differences for nominal data. All tests used were 2 sided, and a P value <.05 was considered significant.

Results

Participants

We enrolled 58 adults who met the study criteria, 20 participants were lost to follow-up during the 10-day baseline period, and 38 participants were randomized to the treatment arm (APP, n=19) and control arm (CON, n=19; Figure 1).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. APP: experimental group (access to artificial intelligence dietary mobile app); CON: control group.



During the intervention period, 13 participants were lost to attrition, and 25 participants completed the study in its totality (CON, n=11; APP, n=14). The COVID-19 pandemic accelerated during the intervention period and accounts for some of the attrition, as stated by those participants who responded to feedback questionnaires. Baseline characteristics at the start of the intervention phase for those completing the study did not differ significantly between the CON and APP groups (Table 1).

Table 1. Baseline characteristics.

Participant characteristics	CON[a] (n=11)	APP[b] (n=14)
Gender, n		
Male	1	2
Female	10	12
Age (years), mean (SD)	25.7 (11.9)	27.2 (9.5)
BMI (kg/m^2), mean (SD)	25.0 (3.3)	27.7 (5.8)
Physical activity (METS/wk[c]), mean (SD)	57.8 (39.3)	60.1 (37.0)
Rome IV, n		
Constipation	4	1
Diarrhea	3	8
Mixed	4	5
IBS[d] symptom score, mean (SD)	275 (56)	272 (43)
Low FODMAP[e] knowledge score, mean (SD)	32.6 (6.5)	29.8 (8.5)
Low FODMAP adherence score, mean (SD)	57.3 (21.8)	65.6 (26.6)
Quality of life score, mean (SD)	117.2 (32.5)	107.4 (27.4)

[a]CON: control group.

[b]APP: experimental group (access to artificial intelligence dietary mobile app).

[c]METS/wk: metabolic equivalents per week.

[d]IBS: irritable bowel syndrome.

[e]FODMAP: fermentable, oligo-, di-, monosaccharide and polyol (diet).

IBS Diagnosis

IBS diagnosis, as measured by the Rome IV criteria, decreased in both groups after the intervention. In the APP group, 6 of 14 participants no longer met the criteria for IBS diagnosis after completion of the intervention. In the CON group, 5 of 11 participants no longer met the criteria for IBS diagnosis at the end of the trial. IBS-D was the most common diagnosis at study completion (CON: IBS-D=2; APP: IBS-D=5); IBS-C and IBS-M were the least diagnosed conditions at study completion (CON: IBS-C=2, IBS-M=2; APP: IBS-C=2, IBS-M=1).

IBS-SSS

The total IBS symptom scores improved over the intervention period for the sample as a whole; however, the change in scores did not differ significantly between groups (CON: –138; APP: –170; *P*=.37; Table 2). Of the 5 individual symptom component scores, the bowel habit dissatisfaction score showed a significant between-group reduction in symptoms. The APP group reported a 2-fold improvement in bowel habit scores in comparison to the CON group (APP: –47; CON: –24; *P*=.05; Figure 2). Significant improvements were noted for the remaining 4 component scores over the 30-day intervention within both groups, but these changes did not vary significantly between groups.

Table 2. Thirty-day change in survey scores.

Survey	CON[a] (n=11), median (IQR)	APP[b] (n=14), median (IQR)	*P* value[c]
IBS[d] symptoms	–123 (–235, –57)	–165 (–238, –116)	.37
Low FODMAP[e] knowledge	10.4 (7.4, 14.0)	8.3 (4.4, 13.1)	.50
FODMAP intake	–6.0 (–27.0, 9.5)	–14.0 (–41.4, 3.3)	.85
Quality of life	7.0 (6.0, 18.0)	21.5 (12.0, 40.3)	.04

[a]CON: control group.

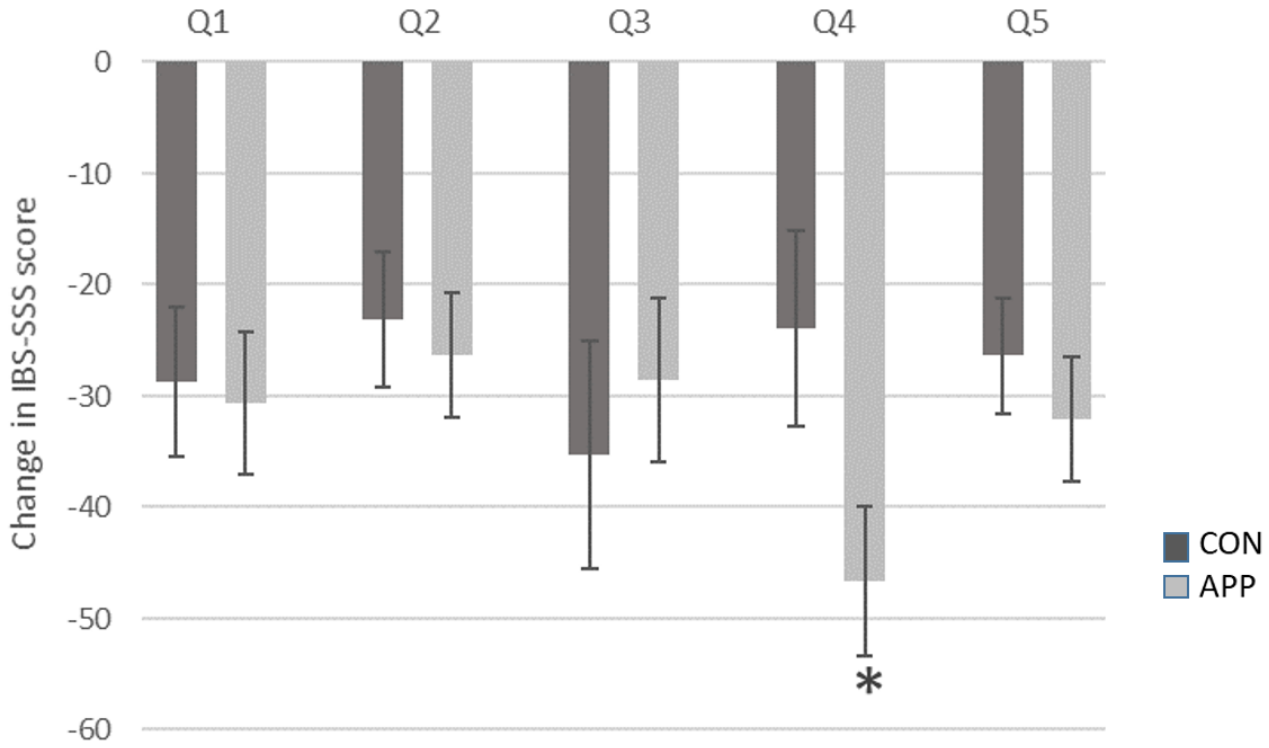
[b]APP: experimental group (access to artificial intelligence dietary mobile app).

[c]*P* values were determined by Mann-Whitney U test.

[d]IBS: irritable bowel syndrome.

[e]FODMAP: fermentable, oligo-, di-, monosaccharide and polyol (diet).

Figure 2. Thirty-day change (mean±SD) in IBS-SSS scores by question. APP: experimental group (access to artificial intelligence dietary mobile app); CON: control group; IBS-SSS: irritable bowel syndrome symptom severity scale; Q1: abdominal pain severity; Q2: abdominal pain frequency; Q3: abdominal pain distension severity; Q4: bowel habit dissatisfaction; Q5: quality of life interference due to the aforementioned symptoms. *Significant difference between CON and APP groups (*P*=.05, Mann-Whitney U test).



Low FODMAP Diet Intake Knowledge

LFD knowledge scores improved over time in both the APP and CON groups, but there was no difference between groups for change in knowledge scores. Adherence to the LFD improved during the study as indicated by a decrease in FODMAP intake scores over the intervention period for both groups. However, the change in intake scores between groups following the intervention period did not differ significantly (Table 2).

Quality of Life

Quality of life scores improved to a greater degree in APP participants compared to CON participants (*P*=.04; Table 2). Furthermore, improvement in quality of life scores was correlated to improvement in IBS symptom scores in the APP group (*r*=–0.598; *P*=.02) but not in the CON group (*r*=–0.183; *P*=.59).

Discussion

Principal Results

These data suggest that supplementing standard IBS dietary education with an AI dietary mobile app that tailors the LFD to specific users' needs improves several health outcomes for individuals with IBS. Specifically, the Heali mobile app helped participants improve their quality of life outcomes and bowel habit symptoms. Quality of life scores rose in both groups; however, the rise in the APP group was 2.6-fold greater than that in the CON group. Poor quality of life is well documented in patients with IBS, and improvement in IBS symptoms is related to improvement in quality of life, such as decreasing the cost of health care, fewer missed days of work, and greater sense of control [37].

Comparison With Prior Work

Similar to the present trial, Kortlever et al [38] recently demonstrated that adoption of the LFD in IBS patients improved GI symptoms and quality of life after 6 weeks of diet adherence. To achieve these results, patients consulted with dietitians at private dietary centers for detailed diet reviews and personalized diet counseling. The data herein suggest that similar outcomes can be achieved with a mobile app, which can thus eliminate certain barriers, such as time commitment, cost, and access to health professionals, making IBS treatment accessible and convenient for a large segment of the population. Although significant quality of life improvements were found in the app group, Pederson et al [10] found that IBS-D participants have a greater response to web-based treatment specifically for quality of life outcomes. In this study, IBS-D participants made up 57% of those in the APP group and 23% of those in the CON group. Although this study was a pilot, future iterations should stratify by IBS subtype to evenly spread this confounding factor.

It is noteworthy that both treatments presented herein improved outcome measures, including diet adherence; yet, mobile app use did demonstrate added benefits. Consumers are increasingly relying on their smartphones for news and information and to conduct their personal business, including managing their health.

Mobile apps that provide ready access to accurate, detailed, and personalized diet information can enable individuals to make lifestyle changes with confidence to improve health. Heali provides evidence-based information and real-time feedback on food choices via a personalized match rating (using the traffic light system) to support user adherence to difficult diets like the LFD.

To our knowledge, Heali is the first mobile app to use AI to support dietary adherence to the LFD and to support treatment of IBS symptoms. This is significant considering that peer-reviewed mobile apps using AI are able to optimize support of chronic disease treatment of conditions such as diabetes [23,39] and hypertension [25]; however, these apps have not been tested on less prominent conditions like IBS. This app-based intervention is especially novel as few studies have explored the possibilities of AI to improve dietary adherence, especially to the LFD [23,26,39]. This cements the novelty of this study on two fronts: (1) the use of AI to improve quality of life in patients with IBS and (2) the use of AI to improve IBS disease outcomes.

Strengths and Limitations

The strengths of this study include the randomization of participants to the intervention and control groups, which allowed dispersion of confounding variables between groups. Further, the entire study, from recruitment to implementation and assessment, was completed online, allowing broad participation without location limitations. The online nature of the study also decreased study costs, as it only required desktop support for implementation and completion.

The limitations of this study include the potential for self-reporting bias, as all surveys were conducted online, while there was also a lack of blinding of the researcher and the participant to the study groups. Participants were randomized via random number generation and not stratified based on severity or type, presenting the potential for bias when comparing between groups. However, participants were not randomized to groups until study day 10, which was after the collection of baseline data. Although anthropometric measures (bodyweight, height) were collected at baseline, they were not collected at the end of the trial, and it is possible that changes in body weight influenced the outcome variables. The change in quality of life was not related to baseline bodyweight or BMI. The small sample size of this pilot trial limits the ability to interpret or generalize findings to other patient populations. Finally, all participants completed the majority of their intervention over the early course of the 2020 COVID-19 pandemic, which reportedly affected participation and retention. Given that this was a dietary restriction study in participants with IBS, it is important to note that household goods (eg, nonperishable foods, proteins, and paper products such as toilet paper and paper towels) were in scarce supply during this time. As a result, it is likely that COVID-19 affected adherence across both groups potentially even more than was stated by participants. The COVID-19 pandemic likely also led to elevated stress, which, in addition to dietary nonadherence, has been shown to exacerbate IBS symptoms [7,8,40,41]. It is therefore important to consider that results reported on symptom severity screeners, progression of the disease, and the quality of life of the participants within this study might have been negatively impacted by the COVID-19 outbreak.

Conclusions

This pilot study provides preliminary evidence that the Heali app may provide therapeutic benefit to its users with IBS. Results showed that the Heali app was able to significantly increase quality of life outcomes in IBS participants over a 30-day intervention period. These findings warrant further research using larger sample sizes. Although this study focused on patients with IBS and the LFD, the variety of additional interventions available via the Heali app suggest possible benefits to individuals with other conditions whose symptoms are attenuated through therapeutic dietary adherence.

Acknowledgments

The authors wish to acknowledge Heali AI for their support in objectively testing the efficacy of their novel application and providing the participants 6 months of free app use after the study; Drs Lucinda Harris, Douglas Drossman, Cecil Rooker, and Johannah Ruddy for their support during the recruitment phase; the International Foundation for Gastrointestinal Disorders for posting our flyer on the clinical trials website; and the Arizona State University Graduate Research Support Program for their funding provided to incentivize participants.

Authors' Contributions

Data were collected by AJR and provided unmodified to CSJ, who independently conducted data analysis and wrote the results section. RH provided consultation and editing support, and all authors reviewed the paper for accuracy.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 5225 KB - jmir_v23i3e24134_app1.pdf]

References

1. Longstreth GF, Thompson WG, Chey WD, Houghton LA, Mearin F, Spiller RC. Functional bowel disorders. Gastroenterology 2006 Apr;130(5):1480-1491. [doi: 10.1053/j.gastro.2005.11.061] [Medline: 16678561]
2. Sperber AD, Bangdiwala SI, Drossman DA, Ghoshal UC, Simren M, Tack J, et al. Worldwide prevalence and burden of functional gastrointestinal disorders, results of Rome Foundation Global Study. Gastroenterology 2021 Jan;160(1):99-114.e3 [FREE Full text] [doi: 10.1053/j.gastro.2020.04.014] [Medline: 32294476]
3. Sandler RS, Everhart JE, Donowitz M, Adams E, Cronin K, Goodman C, et al. The burden of selected digestive diseases in the United States. Gastroenterology 2002 May;122(5):1500-1511. [doi: 10.1053/gast.2002.32978] [Medline: 11984534]
4. Hulisz D. The burden of illness of irritable bowel syndrome: current challenges and hope for the future. J Manag Care Pharm 2004;10(4):299-309. [doi: 10.18553/jmcp.2004.10.4.299] [Medline: 15298528]
5. Nyrop KA, Palsson OS, Levy RL, Von Korff M, Feld AD, Turner MJ, et al. Costs of health care for irritable bowel syndrome, chronic constipation, functional diarrhoea and functional abdominal pain. Aliment Pharmacol Ther 2007 Jul 15;26(2):237-248 [FREE Full text] [doi: 10.1111/j.1365-2036.2007.03370.x] [Medline: 17593069]
6. Peery AF, Crockett SD, Murphy CC, Lund JL, Dellon ES, Williams JL, et al. Burden and cost of gastrointestinal, liver, and pancreatic diseases in the United States: update 2018. Gastroenterology 2019 Dec;156(1):254-272.e11. [doi: 10.1053/j.gastro.2018.08.063] [Medline: 30315778]
7. Ghoshal UC. Marshall and Warren Lecture 2019: A paradigm shift in pathophysiological basis of irritable bowel syndrome and its implication on treatment. J Gastroenterol Hepatol 2020 May;35(5):712-721. [doi: 10.1111/jgh.15032] [Medline: 32162356]
8. Enck P, Mazurak N. The "Biology-First" Hypothesis: functional disorders may begin and end with biology-A scoping review. Neurogastroenterol Motil 2018 Oct;30(10):e13394. [doi: 10.1111/nmo.13394] [Medline: 29956418]
9. Gibson PR, Shepherd SJ. Evidence-based dietary management of functional gastrointestinal symptoms: the FODMAP approach. J Gastroenterol Hepatol 2010 Feb;25(2):252-258. [doi: 10.1111/j.1440-1746.2009.06149.x] [Medline: 20136989]
10. Pedersen N, Andersen NN, Végh Z, Jensen L, Ankersen DV, Felding M, et al. Ehealth: low FODMAP diet vs Lactobacillus rhamnosus GG in irritable bowel syndrome. World J Gastroenterol 2014 Nov 21;20(43):16215-16226 [FREE Full text] [doi: 10.3748/wjg.v20.i43.16215] [Medline: 25473176]
11. Pedersen N. EHealth: self-management in inflammatory bowel disease and in irritable bowel syndrome using novel constant-care web applications. EHealth by constant-care in IBD and IBS. Dan Med J 2015 Dec;62(12):B5168 [FREE Full text] [Medline: 26621403]
12. Pedersen N, Ankersen DV, Felding M, Wachmann H, Végh Z, Molzen L, et al. Low-FODMAP diet reduces irritable bowel symptoms in patients with inflammatory bowel disease. World J Gastroenterol 2017 May 14;23(18):3356-3366 [FREE Full text] [doi: 10.3748/wjg.v23.i18.3356] [Medline: 28566897]
13. Eswaran SL, Chey WD, Han-Markey T, Ball S, Jackson K. A randomized controlled trial comparing the low FODMAP diet vs. modified NICE guidelines in US adults with IBS-D. Am J Gastroenterol 2016 Dec;111(12):1824-1832. [doi: 10.1038/ajg.2016.434] [Medline: 27725652]
14. Vincenzi M, Del Ciondolo I, Pasquini E, Gennai K, Paolini B. Effects of a low FODMAP diet and specific carbohydrate diet on symptoms and nutritional adequacy of patients with irritable bowel syndrome: preliminary results of a single-blinded randomized trial. J Transl Int Med 2017 Jun;5(2):120-126 [FREE Full text] [doi: 10.1515/jtim-2017-0004] [Medline: 28721345]
15. Eswaran S, Chey WD, Jackson K, Pillai S, Chey SW, Han-Markey T. A diet low in fermentable oligo-, di-, and monosaccharides and polyols improves quality of life and reduces activity impairment in patients with irritable bowel syndrome and diarrhea. Clin Gastroenterol Hepatol 2017 Dec;15(12):1890-1899.e3. [doi: 10.1016/j.cgh.2017.06.044] [Medline: 28668539]
16. Staudacher HM, Lomer MCE, Farquharson FM, Louis P, Fava F, Franciosi E, et al. A diet low in FODMPs reduces symptoms in patients with irritable bowel syndrome and a probiotic restores Bifidobacterium species: a randomized controlled trial. Gastroenterology 2017 Oct;153(4):936-947. [doi: 10.1053/j.gastro.2017.06.010] [Medline: 28625832]
17. Altobelli E, Del Negro V, Angeletti PM, Latella G. Low-FODMAP diet improves irritable bowel syndrome symptoms: a meta-analysis. Nutrients 2017 Aug 26;9(9):940 [FREE Full text] [doi: 10.3390/nu9090940] [Medline: 28846594]
18. McIntosh K, Reed DE, Schneider T, Dang F, Keshteli AH, De Palma G, et al. FODMAPs alter symptoms and the metabolome of patients with IBS: a randomised controlled trial. Gut 2017 Jul;66(7):1241-1251. [doi: 10.1136/gutjnl-2015-311339] [Medline: 26976734]
19. Cangemi DJ, Lacy BE. Management of irritable bowel syndrome with diarrhea: a review of nonpharmacological and pharmacological interventions. Therap Adv Gastroenterol 2019;12:1756284819878950 [FREE Full text] [doi: 10.1177/1756284819878950] [Medline: 31632456]
20. Lenhart A, Ferch C, Shaw M, Chey WD. Use of dietary management in irritable bowel syndrome: results of a survey of over 1500 United States gastroenterologists. J Neurogastroenterol Motil 2018 Jul 30;24(3):437-451 [FREE Full text] [doi: 10.5056/jnm17116] [Medline: 29886578]

21. Trott N, Aziz I, Rej A, Surendran Sanders D. How patients with IBS use low FODMAP dietary information provided by general practitioners and gastroenterologists: a qualitative study. Nutrients 2019 Jun 11;11(6):1313 [FREE Full text] [doi: 10.3390/nu11061313] [Medline: 31212668]
22. Morris ME, Aguilera A. Mobile, social, and wearable computing and the evolution of psychological practice. Prof Psychol Res Pr 2012 Dec;43(6):622-626 [FREE Full text] [doi: 10.1037/a0029041] [Medline: 25587207]
23. Stein N, Brooks K. A fully automated conversational artificial intelligence for weight loss: longitudinal observational study among overweight and obese adults. JMIR Diabetes 2017 Nov 01;2(2):e28 [FREE Full text] [doi: 10.2196/diabetes.8590] [Medline: 30291087]
24. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. NPJ Digit Med 2018;1:5 [FREE Full text] [doi: 10.1038/s41746-017-0012-2] [Medline: 31304291]
25. Guthrie NL, Berman MA, Edwards KL, Appelbaum KJ, Dey S, Carpenter J, et al. Achieving rapid blood pressure control with digital therapeutics: retrospective cohort and machine learning study. JMIR Cardio 2019 Mar 12;3(1):e13030 [FREE Full text] [doi: 10.2196/13030] [Medline: 31758792]
26. Guthrie NL, Carpenter J, Edwards KL, Appelbaum KJ, Dey S, Eisenberg DM, et al. Emergence of digital biomarkers to predict and modify treatment efficacy: machine learning study. BMJ Open 2019 Jul 23;9(7):e030710 [FREE Full text] [doi: 10.1136/bmjopen-2019-030710] [Medline: 31337662]
27. Elkjaer M, Burisch J, Avnstrøm S, Lynge E, Munkholm P. Development of a web-based concept for patients with ulcerative colitis and 5-aminosalicylic acid treatment. Eur J Gastroenterol Hepatol 2010 Jun;22(6):695-704. [doi: 10.1097/MEG.0b013e32832e0a18] [Medline: 19543101]
28. Pedersen N, Vegh Z, Burisch J, Jensen L, Ankersen DV, Felding M, et al. Ehealth monitoring in irritable bowel syndrome patients treated with low fermentable oligo-, di-, mono-saccharides and polyols diet. World J Gastroenterol 2014 Jun 07;20(21):6680-6684 [FREE Full text] [doi: 10.3748/wjg.v20.i21.6680] [Medline: 24914395]
29. Palsson OS, Whitehead WE, van Tilburg MAL, Chang L, Chey W, Crowell MD, et al. Development and validation of the Rome IV diagnostic questionnaire for adults. Gastroenterology 2016 May 01;150(6):1481-1491 [FREE Full text] [doi: 10.1053/j.gastro.2016.02.014] [Medline: 27144634]
30. Zahedi MJ, Behrouz V, Azimi M. Low fermentable oligo-di-mono-saccharides and polyols diet versus general dietary advice in patients with diarrhea-predominant irritable bowel syndrome: a randomized controlled trial. J Gastroenterol Hepatol 2018 Jun;33(6):1192-1199. [doi: 10.1111/jgh.14051] [Medline: 29159993]
31. My Nutrition Health. University of Michigan Health System. URL: http://www.myginutrition.com/ [accessed 2020-02-14]
32. Godin G, Shephard RJ. A simple method to assess exercise behavior in the community. Can J Appl Sport Sci 1985 Sep;10(3):141-146. [Medline: 4053261]
33. Francis CY, Morris J, Whorwell PJ. The irritable bowel severity scoring system: a simple method of monitoring irritable bowel syndrome and its progress. Aliment Pharmacol Ther 1997 Apr;11(2):395-402 [FREE Full text] [doi: 10.1046/j.1365-2036.1997.142318000.x] [Medline: 9146781]
34. NHANES food questionnaire. Centers for Disease Control and Prevention. 2018. URL: https://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/tq_fpq_c.pdf [accessed 2020-02-13]
35. Gréa Krause C, Beer-Borst S, Sommerhalder K, Hayoz S, Abel T. A short food literacy questionnaire (SFLQ) for adults: findings from a Swiss validation study. Appetite 2018 Jan 01;120:275-280 [FREE Full text] [doi: 10.1016/j.appet.2017.08.039] [Medline: 28912107]
36. WHOQOL-BREF Introduction, administration, scoring and generic version of the assessment. The World Health Organization. Geneva, Switzerland; 2004. URL: https://www.who.int/mental_health/media/en/76.pdf?ua=1 [accessed 2020-01-11]
37. Böhn L, Störsrud S, Törnblom H, Bengtsson U, Simrén M. Self-reported food-related gastrointestinal symptoms in IBS are common and associated with more severe symptoms and reduced quality of life. Am J Gastroenterol 2013 May;108(5):634-641. [doi: 10.1038/ajg.2013.105] [Medline: 23644955]
38. Kortlever TL, Ten Bokkel Huinink S, Offereins M, Hebblethwaite C, O'Brien L, Leeper J, et al. Low-FODMAP diet is associated with improved quality of life in IBS patients-a prospective observational study. Nutr Clin Pract 2019 Aug;34(4):623-630. [doi: 10.1002/ncp.10233] [Medline: 30644587]
39. Berman MA, Guthrie NL, Edwards KL, Appelbaum KJ, Njike VY, Eisenberg DM, et al. Change in glycemic control with use of a digital therapeutic in adults with type 2 diabetes: cohort study. JMIR Diabetes 2018 Feb 14;3(1):e4 [FREE Full text] [doi: 10.2196/diabetes.9591] [Medline: 30291074]
40. Wang XJ, Camilleri M. Personalized medicine in functional gastrointestinal disorders: understanding pathogenesis to increase diagnostic and treatment efficacy. World J Gastroenterol 2019 Mar 14;25(10):1185-1196 [FREE Full text] [doi: 10.3748/wjg.v25.i10.1185] [Medline: 30886502]
41. Brzozowski B, Mazur-Bialy A, Pajdo R, Kwiecien S, Bilski J, Zwolinska-Wcislo M, et al. Mechanisms by which stress affects the experimental and clinical inflammatory bowel disease (IBD): role of brain-gut axis. Curr Neuropharmacol 2016;14(8):892-900 [FREE Full text] [doi: 10.2174/1570159x14666160404124127] [Medline: 27040468]

Abbreviations

AI: artificial intelligence
APP: experimental group (access to artificial intelligence dietary mobile app)
CON: control group
FODMAPs: fermentable, oligo-, di-, monosaccharides and polyols
GI: gastrointestinal
IBS: irritable bowel syndrome
IBS-C: IBS with constipation
IBS-D: IBS with diarrhea
IBS-M: IBS mixed typed
IBS-SSS: IBS symptom severity scale
IBS-U: IBS unsubtyped
LFD: low FODMAP diet
LFDA: low FODMAP dietary consumption
LFDK: low FODMAP dietary knowledge
WHO: World Health Organization

Edited by G Eysenbach; submitted 04.09.20; peer-reviewed by D Ankersen, M K.; comments to author 12.09.20; revised version received 03.11.20; accepted 15.01.21; published 02.03.21.

Please cite as:
Rafferty AJ, Hall R, Johnston CS
A Novel Mobile App (Heali) for Disease Treatment in Participants With Irritable Bowel Syndrome: Randomized Controlled Pilot Trial
J Med Internet Res 2021;23(3):e24134
URL: https://www.jmir.org/2021/3/e24134
doi:10.2196/24134
PMID:33650977

Original Paper

Development of a Web-Based Mindfulness Program for People With Multiple Sclerosis: Qualitative Co-Design Study

Amy-Lee Sesel[1], MPsych, PhD; Louise Sharpe[1], MPsych, PhD; Heidi N Beadnall[2,3], MBBS, PhD; Michael H Barnett[2,3], MBBS, PhD; Marianna Szabo[1], MPsych, PhD; Sharon L Naismith[1,2], MPsych, DPsych(Neuro)

[1]School of Psychology, University of Sydney, Sydney, Australia

[2]Brain and Mind Centre, University of Sydney, Sydney, Australia

[3]Neurology Department, Royal Prince Alfred Hospital, Sydney, Australia

Corresponding Author:
Amy-Lee Sesel, MPsych, PhD
School of Psychology
University of Sydney
Camperdown NSW
Sydney, 2006
Australia
Phone: 61 2 9351 4558
Email: amy-lee.sesel@sydney.edu.au

Abstract

Background: Mindfulness-based stress reduction is an efficacious treatment for people with chronic health problems; however, it is highly intensive and time-consuming, which is a barrier for service provision.

Objective: This study aims to develop an internet-delivered adapted version of mindfulness-based stress reduction for people with multiple sclerosis to make the intervention more accessible.

Methods: We co-designed a web-based mindfulness program with end users, that is, people with multiple sclerosis (N=19). Iterative feedback was also collected from a subsample of the initial group of end users (n=11), and the program was reviewed by experts (n=8).

Results: We identified three main themes common to people with multiple sclerosis: dealing with uncertainty and fears for the future, grief and loss, and social isolation. These themes were incorporated into narratives throughout the program. People with multiple sclerosis who reviewed the program gave feedback that the program was relatable, feasible, and acceptable. Experts agreed that the program appropriately represented the main tenets of mindfulness. Iterative feedback was used to further refine the program.

Conclusions: The web-based mindfulness program that we developed was viewed positively by both experts and end users. The program reflects common concerns for people with multiple sclerosis and has the potential to meet important unmet psychological needs. A randomized controlled trial was planned to determine the efficacy of the program.

(J Med Internet Res 2021;23(3):e19309) doi:10.2196/19309

KEYWORDS

multiple sclerosis; mindfulness; depression; program development; internet intervention; qualitative research

Introduction

Background

People with multiple sclerosis live with high rates of depression and anxiety, with prevalence rates of approximately 31% [1] and 22% [1,2], respectively. However, there is relatively little literature on psychological treatments for people with multiple sclerosis compared with such treatments for people with other chronic health conditions, and the results of these trials are mixed [3-7].

A recent meta-analysis of 13 studies in multiple sclerosis (MS) [8] revealed that psychosocial interventions significantly improve depression, anxiety, fatigue, and mental and total health-related quality of life (HRQoL). However, effect sizes were small, and for physical HRQoL, the effect was moderated by therapy type. Specifically, although there was no clear benefit of cognitive behavior therapy (CBT) for physical HRQoL, other

psychosocial therapies such as relaxation and mindfulness had a moderate effect (*d*=0.570).

A single study of mindfulness-based stress reduction (MBSR) [9] demonstrated the largest effect sizes of all included studies for depression (*d*=0.8) and anxiety (*d*=0.6). This was a high-quality randomized controlled trial (RCT) of 150 people with multiple sclerosis, showing that face-to-face MBSR was effective across a range of outcomes compared with treatment as usual. Indeed, mindfulness has been shown to have significant mental health benefits for people with a range of physical health conditions, including chronic pain, fibromyalgia, arthritis, and cancer [10,11]. However, the number of trials in the MS literature is limited, and the sample sizes are generally small [3-7]. There is another more recent study of 62 people with multiple sclerosis. This study did not find any significant differences between MBSR and an active control group. However, this study was underpowered [12].

Both RCTs in MS were based on the traditional MBSR program [13] and involved a high dosage of face-to-face intervention (27 hours and 22 hours, respectively), including a full-day workshop. Such intensive treatments are costly, and barriers such as reduced mobility limit accessibility for people with multiple sclerosis. Finding alternative modes of delivery could potentially improve the access, cost-effectiveness, and scalability of interventions. Internet-delivered mindfulness programs have been found to be effective in mental health settings [14] and in chronic disease populations such as cancer, irritable bowel syndrome, and fibromyalgia [15-17]. There is only 1 RCT evaluating the efficacy of online meditation training for people with multiple sclerosis. This study found significant treatment effects for HRQoL, depression, anxiety, and sleep problems at postintervention, but the benefits were not maintained [18]. Although this program was an MS-specific intervention, it was not evident that people with multiple sclerosis were involved in its design. Further research in this field is needed to clarify the mixed results of web-based mindfulness studies and to determine whether a more tailored intervention with consumer input would lead to sustained, longer-term improvements in mental health.

Objectives

The aim of this study is to develop a web-based mindfulness program tailored specifically for people with multiple sclerosis via a qualitative investigation of (1) the psychological experiences and unmet needs of people with multiple sclerosis, (2) attitudes toward a proposed web-based mindfulness program designed to address such needs, and (3) iterative feedback from people with multiple sclerosis and experts in the field.

Methods

Recruitment

Participants were recruited from the MS Clinic, Brain and Mind Centre, Sydney, Australia, between August 2017 and July 2018. Ethics approval was obtained from the University of Sydney Human Research Ethics Committee (2016/049). Potential participants were approached at the MS Clinic waiting room consecutively and given an invitation letter with the opportunity to fill out their contact details. We did not exclude participants with comorbid physical or mental health conditions. Participants were required to be older than 18 years and to have a neurologist-confirmed diagnosis of MS. Interview times were arranged via telephone and were conducted face-to-face at the MS Clinic at an appointed time, with written consent, by a registered psychologist (AS).

Data Collection

Demographic data were collected via interviews and self-report questionnaires. To characterize the sample, we administered valid and reliable measures of depression, anxiety, and pain. Depression symptomatology was measured using the Center for Epidemiological Studies for Depression questionnaire [19] (CES-D; 20-item; range 0-60); anxiety was measured using the State-Trait Anxiety Inventory [20] (STAI) via 2 subscales, assessing state and trait anxiety (both 20-item; range 20-80); and pain was measured using a 10-point visual analogue scale [21]. Semistructured interviews (approximately 50 min in length on average) were recorded and transcribed verbatim. All participants were asked the same semistructured interview questions (Multimedia Appendix 1) with appropriate follow-up questions.

Data Analysis

Data were analyzed using NVivo Qualitative Data Analysis Software [22], via inductive coding at the semantic level, consistent with the Miles and Huberman framework for qualitative data analysis [23]. Analyses were ceased at the point of theme saturation, where 3 transcripts revealed no new themes. Individual transcripts were analyzed and coded independently by 2 authors (AS and LS). The researchers then met with a third researcher (SN) to confirm the themes and coding framework. Any disagreements were resolved by consensus.

Program Development

The web-based mindfulness program was developed on the basis of core components from the MBSR program by Kabat-Zinn [24], with the exception of yoga training. Hatha yoga was omitted from the program because it could not be sufficiently supervised and because the program was designed to target people with multiple sclerosis with varying degrees of mobility. The program was tailored to address the main themes derived from the qualitative interviews with people with MS by integrating case examples from unique, fictional characters that varied in terms of age, gender, race, disease course, and duration. These case examples were used to normalize the day-to-day challenges of people with multiple sclerosis and demonstrate how MBSR could be used to manage MS-related symptoms and improve HRQoL. The program comprised five 15-min web-based modules (Table 1), designed to be delivered over 8 weeks. We opted for 5 modules because there is some evidence that having at least four modules is optimal in other internet interventions (eg, CBT insomnia) [25], but we were unable to provide all meditations in the 4 modules. Furthermore, there is evidence of good adherence to other CBT-based interventions that offer 5 modules over 8 weeks [26,27].

Table 1. Web-based mindfulness program content.

Module	Topics covered	Mindfulness-based stress reduction principles	Formal practice	Informal practice
Module 1: Introduction to mindfulness meditation	Anxiety	• Definition of mindfulness[a] • Benefits of mindfulness • Simple awareness[b]	Awareness of breath meditation	Choose one daily activity to do mindfully
Module 2: Recognizing signs of stress	Stress and fatigue	• Definition of stress • Signs of stress • The body scan[c]	Body scan meditation	Practice brief body scan before a stressful event
Module 3: Dealing with difficult sensations and emotions	Pain	• Cultivating equanimity[d] • Swapping judgment for curiosity • Letting go	Sitting meditation (open awareness)	Cultivate equanimity toward an everyday conflict
Module 4: Dealing with difficult thoughts	Low mood	• Using the breath as an anchor • The 4-step process of nonidentification[e] • The mountain meditation	Mountain meditation	Write down a thought, practice nonidentification, and then write it down again
Module 5: Mindful communication, self-compassion, and relapse prevention	Social isolation	• Active listening and responding (vs reacting) • Cultivating self-compassion • Recap of mindfulness program	Loving-kindness meditation	Complete mindful listening exercise with a partner

[a]Mindfulness: the awareness that arises when we pay attention on purpose, in the present, nonjudgmentally.

[b]Simple awareness: observing what is happening by focusing your attention completely on one thing at a time, without judging your experience in any way.

[c]Body scan: a type of mindfulness meditation that involves concentrating one's attention on one's body, moving from the tip of the toes to the top of the head in a systematic, mindful way.

[d]Equanimity: a neutral state of mind toward all experiences, pleasant and unpleasant, regardless of whether they bring pleasure, joy, or misery.

[e]Nonidentification: a technique that involves being conscious of not becoming attached to whatever thoughts arise within your internal experience, taking a step back, and viewing each thought as an impersonal mental event.

The 5 modules were scheduled as follows: module 1 at the start of week 1 and module 2 at the start of week 2. The remaining modules contain instructions for the next 2 weeks, that is, module 3 would be delivered at the start of week 3 and again for the option of repetition, at the beginning of week 4, etc. The amount of time (ie, 1 or 2 weeks) dedicated to each module was decided based on the complexity of the mindfulness concepts and strategies described. It was written by a registered psychologist (AS) with clinical psychology and mindfulness training in collaboration with a clinical psychologist with experience in the development and evaluation of evidence-based psychological interventions for people with chronic health conditions (LS) and a clinical neuropsychologist with experience in the delivery of web-based therapeutic interventions (SN).

Direct, face-to-face feedback on the program was provided by a subsample of the original participants interviewed (n=11), who indicated that they were interested and willing to review the modules of the program. In total, 8 women and 3 men with MS (10/11 relapsing-remitting; age range: 23-63 years) took part in the review process. Two neurologists and 4 mindfulness experts including 2 clinical psychologists, 1 emergency care physician and 1 general practitioner provided written feedback on the program. Feedback was then coded and analyzed inductively, forming subheadings with overarching themes. Feedback from both people with multiple sclerosis and experts was discussed in team meetings among the 3 lead authors (AS, LS, and SN). Further iterations of the program were cocreated where there was consensus and scope to make the suggested changes.

Results

Demographics

All participants had a neurologist-confirmed diagnosis of MS (N=19). There were 6 males and 13 females. A total of 63% (12/19) of participants had clinically significant depressive symptoms according to a cut-off score of 16 on the CES-D [28]. Moreover, 47% (9/19) of participants had clinically significant levels of state anxiety, according to a cut-off score of 41, and 42% (8/19) of participants had clinically significant levels of trait anxiety, according to a cut-off score of 44 on the STAI [29]. Descriptive statistics are presented in Table 2.

Table 2. Participant demographics (N=19).

Characteristics	Values
Females, n (%)	13 (68)
Age (years), mean (SD)	40.42 (16)
Employment status, n (%)	
Working full-time or part-time	8 (42)
Unemployed	6 (32)
Retired	1 (5)
University student	3 (16)
High school student	1 (5)
Type of MS[a], n (%)	
Primary progressive MS	5 (26)
Relapsing-remitting MS	14 (74)
Walking ability, n (%)	
Wheelchair or scooter	3 (16)
Ambulatory	16 (84)
Psychosocial descriptors	
Pain now, n (%)	8 (42)
Pain, mean (SD)[b]	4.48 (4)
Depressive symptoms, mean (SD)	19.89 (13)
State anxiety, mean (SD)	39.16 (15)
Trait anxiety, mean (SD)	43.89 (14)

[a]MS: multiple sclerosis.

[b]On the basis of the 8 participants who reported experiencing pain.

Thematic Analysis

The thematic analysis of this qualitative study was divided into 3 parts, in accordance with the research aims. For detailed thematic maps, including illustrative example quotations, please see Tables 3-5.

Table 3. Part 1: Psychological experiences and unmet needs of people with multiple sclerosis.

Overarching theme and subtheme	Participant quote
Uncertainty	
When another relapse will occur	"...you really don't know when you're going to have another episode...it's a pretty stressful thing." (P6)
Fear of progression	"I guess I don't want to face the fact that yeah, something could happen and when's it going to happen? I don't know. Nobody's knows so I guess." (P17)
Potential burden on family	"I suppose my biggest worry at the moment is, how long will I stay, or how could I possibly get a little bit better so I can get more strength so I can actually um, perform better in the workplace...cause my wife would be never the person who could go out and work...I don't know what would happen to her." (P1)
Neurologist appointments	"Anytime I have to come to talk to Dr [name] about something or get results...it just destroys me for a week or two." (P2)
Gaps in medical knowledge	"I remember when I first got diagnosed and I asked, what causes this, they said...its multi-factorial, which means we don't know...and you get that answer a lot." (P8)
What is MS[a]-related	"Sometimes when I'm fatigued, or having low mood, it's hard to know whether that's just a normal thing, or whether it is MS influenced." (P6)
Grief and loss	
Loss of future plans	"I'm just feeling really heartbroken about having MS...Our whole future and life just completely changed." (P2)
Difficulty with acceptance	"There's an element of why me, poor me...why is this happening, this can't be happening...and then cry a lot of tears." (P3)
Loss of independence	"I can't do anything that I used to be able to do...I can't play netball, I can't play touch footy, I can't walk in the dark...you lose everything..." (P4)
Loss of cognitive abilities	"...I've missed out on a lot of things in life that you would normally take for granted...like as your young children grow up...I don't remember those years of them..." (P1)
Loss of confidence	"I don't feel like I can rely on my own devices...I wouldn't have the confidence to go and do something on my own that's out of the norm, out of my comfort zone, so that's my thing, is I just feel vulnerable." (P18)
Loss of sense of self and identity	"I can't think normally anymore, and I'm not as quick as I used to be either. Everything's slowly deteriorating, including me and my thoughts." (P7)
Social isolation	
Emotional avoidance	"Well, it's mainly my family, who will sometimes bring it up in conversations...but yeah I really avoid it..." (P3)
Inappropriate responses of family or friends	"They try to help, but like I've had people burst into tears when I tell them, just like really inappropriate because they don't know how to handle it..." (P2)
Lack of understanding	"Nobody really understands, and I don't really have anybody to talk to about it." (P17)
Apathy from others	"I've been brave enough to step out there and say this is what's going on for me, and kind of reaching out a hand for some kind of support or acknowledgement, and getting nothing..." (P15)
Rejection due to stigma	"Yeah, [friends] walked away. Cause I got MS you know...they could catch it!" (P4)
Availability of support	
Lack of follow-up	"Well, I suppose for someone like me I would've appreciated some sort of supportive type of stuff, but there was nothing." (P13)
Lack of funding or resources	"From my experience with the MS Society, every time I ever went to them, they always said they ran out of money and couldn't help me." (P1)
Not knowing how to access support	"...I didn't seek out specific psychological treatment for MS...but I don't know where I would have gone if I did want that." (P8)
Lack of therapists with MS knowledge	"I haven't found anybody [therapist] who specializes in MS." (P2)
Group-based support	
Avoiding others with more progressive disability	"...if I see somebody very disabled by their MS, it's really confronting and upsetting to me, because well, I'm lucky that it's not me now, but there's no way of telling what it will be like in the future, and that's really scary." (P14)

Overarching theme and subtheme	Participant quote
Difficulties with relatability	"I don't really want to talk to people my mother's age with MS." (P19)

[a]MS: multiple sclerosis.

Table 4. Part 2: Attitudes toward the web-based mindfulness program.

Overarching theme and subtheme	Participant quote
Seeking support online	
Anonymity	"I'd probably feel ok because I can do it in the privacy of my own space so it wouldn't draw attention to me." (P3)
Flexibility	"...when I am overtired, when I am hot, when there are issues, I do need that flexibility." (P14)
No need for travel	"I don't think they [online therapies] would replace in-person, but I think they could augment, or be useful in between. And maybe some people would find it really useful if they don't have access to something local..." (P2)
Internet security	"I have big phobias about Internet security and things like that." (P4)
Vision problems	"I can't see..." (P5)
Mindfulness meditation and barriers to practice	
Potential efficacy	"I do have an understanding and I do believe it works...it's very helpful to bring you back into reality, to stop your mind going crazy." (P3)
Difficulties with finding time	"I have had trouble making sure I do it every day, like, finding that time and being strict with myself." (P17)
Difficulties concentrating	"I get distracted really easily..." (P14)
Low mood	"I think it's a depression. And I know if I were to do those things [to improve my mental health] the depression would get better." (P2)
Lack of quiet space	"If I go in my room and close the door, my nieces would just run up and start banging on the door shouting 'Uncle, uncle, uncle'." (P7)
Anticipated program use, preferences, and suggestions	
Interest in participating	"I like the idea of the program, and I think there are lots of people who it would really help." (P6)
Length of meditation	"I would think minimum 20 minutes because I think anything less than that you haven't really got into it, to me, takes a while to get into the zone." (P11)
Case examples	"...if I had a case to go by, I could say yeah ok that sounds familiar, and then go from there...because with a lot of the stuff it's like is this supposed to happen? Does it happen to people with MS[a] or does it happen to everyone? So it would be good to have case studies." (P3)
Email reminders	"I'd be more likely to actually use the reminder if it was e-mail, because then I can sort of flag it and keep track of it." (P10)
Weekly telephone calls	"To me, I would like follow-up contact. Especially if you develop a rapport with somebody, it's good to have follow-up contact. That definitely helps with learning a concept, to get a certain amount of rapport and follow-up." (P1)
More likely to participate if recommended	"I think it would make it less likely for me to get benefit from it because I'd be so skeptical. If one of my doctors said, or if my psychiatrist who had MS said, [to do it] I certainly would." (P8)
More likely to participate if tailored to MS	"...if it's something like 'Yeah we can help to work on that anxiety, or work on some things' and its geared towards people with MS, I think that would be really helpful." (P2)

[a]MS: multiple sclerosis.

Table 5. Part 3: Iterative feedback from people with multiple sclerosis and experts in the field.

Overarching theme and subtheme	Participant quote
Relatability	
Relevance of case examples	"Was it from these interviews that you got the stories for your modules and helped plan the program? I ask this as the stories seem very true to life indeed." (Exp1)[a]
Identification with program characters	"I think it's great. I can relate to Jen and see myself embedded in these slides." (P11)
Language: too simplistic?	"My impression was that it was speaking to children in terms of the graphics and language like reading them a storybook. Is that what you intended?" (Exp2)[b]
Language: is acceptable	"I personally have no issue with it. I feel like that's just kind of the 'mode' you're working in." (P8)
Graphics: too childish?	"Yes, I also thought the images looked childish." (P2)
Graphics: are acceptable	"I personally really like it, because I think the cartoon-style characters get you completely away from stereotypes and to focus on the messages and the content." (P15)
Acceptability of program content	
Analogies	"Yeah, really good, I think the analogies are good, you know, the monkey chatter in the mind." (P17)
Mindfulness principles	"I love the content, the flow and the style." (P15)
Mindfulness training	"I think the text that will be spoken is pretty good and I think follows most mainstream advice about mindfulness and how to teach it." (Exp3)[c]
Include expert	"Having a mindfulness expert as a supervisor is important." (Exp4)[d]
Include scientific evidence	"… the presentation of some of the scientific evidence underpinning mindfulness generally improves acceptance." (Exp3)

[a]Exp1: Expert 1, neurologist, specialty in multiple sclerosis

[b]Exp2: Expert 2, general practitioner with expertise in the delivery of web-based mindfulness interventions

[c]Exp 3: Expert 3, emergency physician with expertise teaching mindfulness to people with multiple sclerosis

[d]Exp 4: Expert 4, clinical psychologist with mindfulness teacher training.

Part 1: The Psychological Experiences and Unmet Needs of People With Multiple Sclerosis

Uncertainty

Most participants reported that their psychological experiences were characterized by uncertainty, worry about the future and possible disease progression, and difficulty grappling with unknowns surrounding the disease. The inability to predict possible relapse was a major source of continual stress and anxiety, and it interfered with planning, for example, whether and when they should have children, how long their physical capabilities would allow them to stay in the workforce, and the potential financial burden that they would place on their families. Going to neurologist appointments and having to face the possibility of receiving bad news, such as the identification of new lesions in the brain or spinal cord, or changing treatments, was a source of excessive worry and distress. Some participants worried about whether their physical or cognitive problems were part of their MS or due to other factors, such as the natural process of aging.

Grief and Loss

An overarching theme was the participants' experience of grief and loss due to the gradual decline in their physical and cognitive abilities and their sense of independence and identity. They also expressed anguish and sorrow over the loss of their illness-free life. Many people with multiple sclerosis reported that they were faced with having to give up on future plans and accept changes to their ability to travel and participate in family life. This process of adjustment experienced by people with multiple sclerosis was depicted as challenging and repetitive, as people developed new or worsening symptoms over time. People with multiple sclerosis experiencing cognitive difficulties described experiences of memory loss and how this impacted their daily functioning as well as their ability to recall cherished memories (eg, of their children growing up), which reinforced feelings of disconnection from family life. Those experiencing physical challenges, such as ataxia, muscle weakness, and bladder dysfunction, reported feelings of helplessness and a great sense of vulnerability, as they gradually lost confidence in themselves. Participants across all ages, genders, and types of MS emphasized the need to separate themselves from the disease, to try to lead a *normal life* and preserve their sense of self and identity, which many feared was at risk of erosion.

Social Isolation

Participants reported experiencing social isolation and difficulties communicating with people in their support networks about their MS. Barriers to receiving social support included difficulties with participating in social activities due to loss of function and emotional avoidance when faced with opening up to people about their MS as well as other people's failure to respond appropriately. Some participants reported that they were given unsolicited and often ill-informed advice from

friends and family, which was perceived as highly distressing and alienating. Many participants reflected a sense of apathy and disinterest on behalf of their wider social circles and work environment and others' general lack of understanding that MS is largely an invisible disease. Two participants reported facing blatant rejection from others because of stigma and unfounded beliefs about MS being contagious.

Availability of Psychological Support

Most participants reported an unmet need for psychological support, which they attributed to a lack of follow-up, lack of funding or resources, lack of knowledge regarding where to go for psychological support, and lack of psychologists or therapists specializing in MS. Of the 63% (12/19) participants with clinically significant symptoms of depression, only half reported that they were either currently seeing a psychologist or had seen one in the past.

Group-Based Psychosocial Support

A major barrier to receiving psychological support was that the overwhelming majority of participants did not want to meet other people with multiple sclerosis in a face-to-face therapeutic setting, despite wanting to know how others cope. This was largely because meeting people with more severe symptomatology or disability generated feelings of uncertainty and fear around their own illness progression. For representative participant quotations for each subtheme, please see Table 3.

These main themes underlying the psychological experiences of people with multiple sclerosis were incorporated into the program development through the use of fictional characters featured in each module, whose stories and case examples reflected the common challenges, and mental-health related difficulties that participants reportedly faced. For screenshot examples, please see Multimedia Appendix 2.

Part 2: Attitudes Toward a Proposed Web-Based Mindfulness-Based Program

Attitudes Toward Internet Use

Participants reported positive attitudes toward seeking psychological support online. It is widely regarded as convenient, easy, and allowing access to therapeutic support with a sense of anonymity, in the privacy and comfort of their homes, without the need for travel. Reservations toward the use of an internet-delivered psychological intervention were noted by 11% (2/19) participants. One experienced MS-related vision problems and another expressed apprehension surrounding internet security.

Attitudes Toward Mindfulness

Participants' experiences and knowledge of mindfulness varied from having no understanding of what it was to having years of experience. Potential barriers to participating in the proposed mindfulness program included not having enough time in the day to meditate, difficulties getting in a routine, problems with attention and concentration, low mood, and inability to find a quiet space to meditate, particularly for those with young families.

Anticipated Use, Program Preferences, and Suggestions

Most participants reported that they would be interested in participating in the proposed web-based mindfulness intervention (see *Methods* section). People with multiple sclerosis with more experience with mindfulness meditation tended to advocate for longer meditation practices. Features that were identified to engage participants in the program and reduce dropout included incorporating case examples, email reminders, and guiding participants through the program via weekly telephone calls. Three participants reported that they would be more likely to participate in the program if their neurologist or psychiatrist recommended it. Many reported that a program tailored to people with multiple sclerosis was preferable compared with a generic web-based mindfulness program. For representative participant quotations, please see Table 4.

Part 3: Iterative Feedback From People With Multiple Sclerosis and Experts in the Field

Relatability of the Program

Most people with multiple sclerosis found the program highly relatable to their experiences of living with MS. In total, 5 participants and 1 neurologist gave specific positive feedback about the case examples. Reservations about the relatability of the program were expressed by 1 expert and 1 participant (who held a PhD) who commented that the program was too simplistic and could be perceived as patronizing. Following this feedback, participants and experts were explicitly asked about their views in this regard, and these concerns were not widely shared by other respondents. Some changes to the program graphics and language were incorporated to further improve the relatability of the program: graphics were changed to make the characters look older and the scenes more realistic, and some of the language was edited to be less repetitive and more direct.

Acceptability of Program Content

In terms of the program content, 8 participants gave specific, positive feedback on the analogies used to explain mindfulness concepts (eg, *monkey chatter*) and how the mindfulness concepts were presented. Experts in mindfulness commented that the concepts were well explained and consistent with MBSR principles; however, some suggested involving a mindfulness expert, which we did. One area in which there was disagreement among experts was on the degree to which more reference to research should be included. Some experts recommended integrating scientific evidence on the efficacy of mindfulness into the program to improve treatment adherence, whereas others disagreed. We decided that including specific scientific research and results of studies was not necessary. For representative quotations of the feedback provided by participants and experts in the field, please see Table 5.

Discussion

Principal Findings

Participants reported psychological support as a common unmet need and were generally positive about the idea of an MS-specific, web-based mindfulness intervention. Among the

participants, the anticipated use of such a program was high, particularly if it was endorsed by treating physicians. A thematic analysis of 19 face-to-face interviews with people with multiple sclerosis revealed 3 overarching themes characterizing their psychological experiences: uncertainty, grief and loss, and social isolation. These themes were used in the development of narratives throughout the mindfulness program.

The first theme of uncertainty is unsurprising, considering that MS is an unpredictable illness. Participants indicated that they were fearful of progression, worried about whether symptoms might indicate deterioration, and worried about relapse. The lack of knowledge about the causes of MS exacerbated the uncertainty and worry and made attending medical appointments the source of anxiety. For some, the fear of progression was related to fear of becoming a burden on their family. Fears of progression have been previously documented in MS and found to be associated with poorer psychosocial outcomes [30]; therefore, we featured these throughout the program.

The second theme was grief and loss. Participants described the struggle to accept the limitations imposed on them by MS. They experienced losses of their future plans, which led to a loss of independence. Specific losses such as loss of cognitive abilities and loss of confidence led to a loss of a sense of self and identity. The importance of loss and its impact on self and identity have been previously described as important in the context of adjustment to illness [31]. Hence, we represented these struggles in the narratives developed for the program.

The third theme was social isolation. In terms of social isolation, people with MS felt that family and friends could not understand their difficulties, which sometimes led to inappropriate responses or apathy from others. Some people with MS described having lost friendships, which they experienced as rejection due to the stigma of having MS. Many people with MS described emotional avoidance, which allowed them to not express their emotions to others as a way of trying to avoid negative responses from others. However, this led to an overwhelming feeling of social isolation, which has been previously recognized in the literature [32].

We used these qualitative results to provide narratives to supplement the mindfulness content of our program. Once a draft version of the program was available on the basis of early interviews, we also gained feedback from both the people with multiple sclerosis and experts in MS and mindfulness. Iterative feedback allowed modifications to be made to the program content and graphics to improve relatability and ensure adherence to MBSR principles. However, overall feedback from mindfulness experts, neurologists, and people with multiple sclerosis were positive. The adapted mindfulness program appeared to be an acceptable and highly relatable program, tailored to the unmet needs and experiences of people with multiple sclerosis, and consistent with the basic principles and teachings of MBSR by Kabat Zinn [24].

As previously described, the primary reason for including the psychological experiences of people with multiple sclerosis was to ensure that the narratives included in our program were relevant to the end users of the program. It seems that the issues raised in our sample were similar to those previously raised in the literature. Prior research has shown that there is a lack of psychological support and continuity of psychosocial care after receiving the initial diagnosis [33,34]. Previous studies have highlighted the difficulties in accepting uncertainty surrounding the diagnosis, treatment, and prognosis of MS [35,36]. Indeed, *illness uncertainty* has been found to predict adjustment problems for people with multiple sclerosis, over and above demographic and disease characteristics [37], and the experience of fears of relapse or progression are well documented. Experiences of grief and loss [38] as well as difficulties in maintaining one's sense of self in the face of physical and cognitive decline have been reported in the literature [31,39,40]. Furthermore, beliefs of people with multiple sclerosis about their illness and illness identity have been found to play a significant role in their psychological adjustment [41]. Previous research suggests that stereotypical images of a person in a wheelchair are regarded as negative symbols of loss, disability, and death for people with multiple sclerosis [42]. This may provide a possible explanation as to why the majority of participants in this study expressed disinterest in participating in group-based psychological interventions or support groups, as participants said they particularly wanted to avoid seeing people with more severe disease than themselves. Finally, it is well known that the psychological well-being of people with multiple sclerosis is worsened by social isolation and difficulties in communicating effectively about the disease with family and friends. In many cases, these feelings of abandonment, isolation, and social withdrawal are reported years after the initial diagnosis [40,43,44]. Commonalities between the themes identified from the people with multiple sclerosis in our study and those reported in the literature indicate that this innovative web-based mindfulness program is likely to be relatable to the target population.

Given the difficulties with anxiety and uncertainty about the future and the experiences of grief and loss experienced by people with multiple sclerosis, a mindfulness-based approach that focuses on the *present moment*, cultivating *acceptance* and *letting go*, seems an appropriate intervention for reducing psychological distress and improving HRQoL. Furthermore, the delivery of a mindfulness intervention, tailored to the needs and experiences of people with multiple sclerosis via the web, can be used to normalize the challenges of living with the disease and decrease feelings of isolation while maintaining anonymity and increasing access to psychological support.

Limitations

Although we attempted to interview a range of people with multiple sclerosis with various demographic and disease characteristics, we acknowledge that the experiences of people with multiple sclerosis reported here were a reflection of a small sample from one recruitment center. Another limitation of this study was that the reviewers were shown PowerPoint versions of the modules and not the web-based multimedia versions, which would have allowed for further analysis of the program's usability. Finally, the evaluation of efficacy was beyond the scope of this study. A study protocol for an RCT evaluating the efficacy of the mindfulness program described in this study has been developed [45]. Clearly, the results of an adequately

powered RCT are needed to determine the efficacy of the program.

Conclusions

In summary, we have developed a web-based mindfulness intervention using an iterative co-design process that reflects the common challenges reported by people with multiple sclerosis. We included narratives that were developed from the themes derived from a qualitative study. The modules appropriately reflected the principles of MBSR, and the included narratives were viewed as relatable by people with multiple sclerosis. Future research will determine whether the developed intervention is effective. If it proves to be beneficial, the developed program would likely meet an important need for people with multiple sclerosis.

Acknowledgments

The authors would like to acknowledge the people with multiple sclerosis who contributed their valuable time and told their stories to aid the development of the mindfulness program described in this study. The authors also thank the experts who offered their generous advice, suggestions, and constructive feedback, including Alice Shires, Professor George Jelinek, and Associate Professor Craig Hassed. This project was supported by a postgraduate scholarship from the National Health and Medical Research Council and Multiple Sclerosis Research Australia. It was part of a competitive application, and these funding bodies made no contribution to the study design, data analysis, interpretation, or writing of this manuscript.

Authors' Contributions

AS conducted all the interviews with people with multiple sclerosis, coded the data into themes, and wrote the original manuscript as well as the mindfulness program. LS contributed to the conceptualization and design of the program, contributed to the conceptualization and coding of main themes, provided feedback on the program, and assisted with drafts of the manuscript. SN contributed to the design of the program, provided feedback on the program, and edited drafts of the manuscript. MS, HB, and MB provided feedback on the program and edited drafts of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview questions.
[DOCX File , 16 KB - jmir_v23i3e19309_app1.docx]

Multimedia Appendix 2
Screenshots from the program, as shown to participants.
[PDF File (Adobe PDF File), 200 KB - jmir_v23i3e19309_app2.pdf]

References

1. Boeschoten RE, Braamse AM, Beekman AT, Cuijpers P, van Oppen P, Dekker J, et al. Prevalence of depression and anxiety in Multiple Sclerosis: A systematic review and meta-analysis. J Neurol Sci 2017 Jan 15;372:331-341. [doi: 10.1016/j.jns.2016.11.067] [Medline: 28017241]
2. Marrie RA, Reingold S, Cohen J, Stuve O, Trojano M, Sorensen PS, et al. The incidence and prevalence of psychiatric disorders in multiple sclerosis: a systematic review. Mult Scler 2015 Mar 12;21(3):305-317 [FREE Full text] [doi: 10.1177/1352458514564487] [Medline: 25583845]
3. Kuspinar A, Rodriguez AM, Mayo NE. The effects of clinical interventions on health-related quality of life in multiple sclerosis: a meta-analysis. Mult Scler 2012 Dec 27;18(12):1686-1704. [doi: 10.1177/1352458512445201] [Medline: 23235779]
4. Hind D, Cotter J, Thake A, Bradburn M, Cooper C, Isaac C, et al. Cognitive behavioural therapy for the treatment of depression in people with multiple sclerosis: a systematic review and meta-analysis. BMC Psychiatry 2014 Jan 9;14(1). [doi: 10.1186/1471-244x-14-5]
5. Fiest K, Walker J, Bernstein C, Graff L, Zarychanski R, Abou-Setta A, et al. Systematic review and meta-analysis of interventions for depression and anxiety in persons with multiple sclerosis. Multiple Sclerosis and Related Disorders 2016 Jan;5:12-26. [doi: 10.1016/j.msard.2015.10.004]
6. Simpson R, Booth J, Lawrence M, Byrne S, Mair F, Mercer S. Mindfulness based interventions in multiple sclerosis - a systematic review. BMC Neurol 2014 Jan 17;14(1). [doi: 10.1186/1471-2377-14-15]
7. Mohr DC, Goodkin DE. Treatment of depression in multiple sclerosis: review and meta-analysis. Clinical Psychology: Science and Practice 1999;6(1):1-9. [doi: 10.1093/clipsy.6.1.1]

8. Sesel A, Sharpe L, Naismith S. Efficacy of Psychosocial Interventions for People with Multiple Sclerosis: A Meta-Analysis of Specific Treatment Effects. Psychother Psychosom 2018 Mar 8;87(2):105-111. [doi: 10.1159/000486806] [Medline: 29518781]
9. Grossman P, Kappos L, Gensicke H, D'Souza M, Mohr DC, Penner IK, et al. MS quality of life, depression, and fatigue improve after mindfulness training: a randomized trial. Neurology 2010 Sep 28;75(13):1141-1149 [FREE Full text] [doi: 10.1212/WNL.0b013e3181f4d80d] [Medline: 20876468]
10. Hofmann SG, Sawyer AT, Witt AA, Oh D. The effect of mindfulness-based therapy on anxiety and depression: A meta-analytic review. J Consult Clin Psychol 2010 Apr;78(2):169-183 [FREE Full text] [doi: 10.1037/a0018555] [Medline: 20350028]
11. Grossman P, Niemann L, Schmidt S, Walach H. Mindfulness-based stress reduction and health benefits. A meta-analysis. J Psychosom Res 2004 Jul;57(1):35-43. [doi: 10.1016/S0022-3999(03)00573-7] [Medline: 15256293]
12. Senders A, Hanes D, Bourdette D, Carson K, Marshall LM, Shinto L. Impact of mindfulness-based stress reduction for people with multiple sclerosis at 8 weeks and 12 months: A randomized clinical trial. Mult Scler 2019 Jul 09;25(8):1178-1188 [FREE Full text] [doi: 10.1177/1352458518786650] [Medline: 29985095]
13. Kabat-Zinn J. An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: theoretical considerations and preliminary results. Gen Hosp Psychiatry 1982 Apr;4(1):33-47. [doi: 10.1016/0163-8343(82)90026-3] [Medline: 7042457]
14. Spijkerman M, Pots W, Bohlmeijer E. Effectiveness of online mindfulness-based interventions in improving mental health: A review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
15. Zernicke KA, Campbell TS, Speca M, McCabe-Ruff K, Flowers S, Carlson LE. A randomized wait-list controlled trial of feasibility and efficacy of an online mindfulness-based cancer recovery program: the eTherapy for cancer applying mindfulness trial. Psychosom Med 2014 May;76(4):257-267. [doi: 10.1097/PSY.0000000000000053] [Medline: 24804884]
16. Ljótsson B, Falk L, Vesterlund AW, Hedman E, Lindfors P, Rück C, et al. Internet-delivered exposure and mindfulness based therapy for irritable bowel syndrome--a randomized controlled trial. Behav Res Ther 2010 Jun;48(6):531-539. [doi: 10.1016/j.brat.2010.03.003] [Medline: 20362976]
17. Davis MC, Zautra AJ. An online mindfulness intervention targeting socioemotional regulation in fibromyalgia: results of a randomized controlled trial. Ann Behav Med 2013 Dec 14;46(3):273-284. [doi: 10.1007/s12160-013-9513-7] [Medline: 23670111]
18. Cavalera C, Rovaris M, Mendozzi L, Pugnetti L, Garegnani M, Castelnuovo G, et al. Online meditation training for people with multiple sclerosis: A randomized controlled trial. Mult Scler 2019 Apr 27;25(4):610-617. [doi: 10.1177/1352458518761187] [Medline: 29485319]
19. Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. Applied Psychological Measurement 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]
20. Spielberger CD, Gorsuch RL, Lushene R, Vagg PR, Jacobs GA. Manual for the state-trait anxiety scale (form Y self-evaluation questionnaire). Palo Alto. CA: Consulting psychologists press; 1983.
21. Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. Pain 1983 Sep;17(1):45-56. [doi: 10.1016/0304-3959(83)90126-4] [Medline: 6226917]
22. Bazeley P, Jackson K. Qualitative data analysis with Nvivo. In: Seaman J, editor. QSR International Pty Ltd. Cambridge, MA. London: Sage Publications; 2013.
23. Miles MB, Huberman AM. Qualitative data analysis: An expanded sourcebook. Thousand Oaks, London, New Delhi: Sage publications; 1994.
24. Kabat-Zinn J. Full catastrophe living, revised edition: how to cope with stress, pain and illness using mindfulness meditation. United Kingdon: Hachette; 2013.
25. Edinger JD, Wohlgemuth WK, Radtke RA, Coffman CJ, Carney CE. Dose-response effects of cognitive-behavioral insomnia therapy: a randomized clinical trial. Sleep 2007;30(2):203-212. [doi: 10.1093/sleep/30.2.203] [Medline: 17326546]
26. Read J, Sharpe L, Burton AL, Arean PA, Raue PJ, McDonald S, et al. A randomized controlled trial of internet-delivered cognitive behaviour therapy to prevent the development of depressive disorders in older adults with multimorbidity. Journal of affective disorders 2020 Mar 01;264:464-473. [doi: 10.1016/j.jad.2019.11.077] [Medline: 31767215]
27. Dear BF, Gandy M, Karin E, Staples LG, Johnston L, Fogliati VJ, et al. The Pain Course: a randomised controlled trial examining an internet-delivered pain management program when provided with different levels of clinician support. Pain 2015 Oct;156(10):1920-1935 [FREE Full text] [doi: 10.1097/j.pain.0000000000000251] [Medline: 26039902]
28. Weissman MM, Sholomskas D, Pottenger M, Prusoff BA, Locke BZ. Assessing depressive symptoms in five psychiatric populations: a validation study. Am J Epidemiol 1977 Sep;106(3):203-214. [doi: 10.1093/oxfordjournals.aje.a112455] [Medline: 900119]
29. Ercan I. Examining Cut-Off Values for the State-Trait Anxiety Inventory/ Examinando los puntajes de corte para el inventario de ansiedad estado-rasgo. Revista Argentina de Clinica Psicologica 2015;24(II):143.

30. Khatibi A, Moradi N, Rahbari N, Salehi T, Dehghani M. Development and Validation of Fear of Relapse Scale for Relapsing-Remitting Multiple Sclerosis: Understanding Stressors in Patients. Front Psychiatry 2020 Mar 20;11:226 [FREE Full text] [doi: 10.3389/fpsyt.2020.00226] [Medline: 32265764]
31. Sharpe L, Curran L. Understanding the process of adjustment to illness. Soc Sci Med 2006 Mar;62(5):1153-1166. [doi: 10.1016/j.socscimed.2005.07.010] [Medline: 16112783]
32. Freeman J, Gorst T, Gunn H, Robens S. "A non-person to the rest of the world": experiences of social isolation amongst severely impaired people with multiple sclerosis. Disabil Rehabil 2020 Aug 18;42(16):2295-2303. [doi: 10.1080/09638288.2018.1557267] [Medline: 30657717]
33. Solari A, Acquarone N, Pucci E, Martinelli V, Marrosu M, Trojano M, et al. Communicating the diagnosis of multiple sclerosis - a qualitative study. Mult Scler 2007 Jul 09;13(6):763-769. [doi: 10.1177/1352458506074689] [Medline: 17613605]
34. Bogosian A, Morgan M, Moss-Morris R. Multiple challenges for people after transitioning to secondary progressive multiple sclerosis: a qualitative study. BMJ Open 2019 Mar 08;9(3):e026421 [FREE Full text] [doi: 10.1136/bmjopen-2018-026421] [Medline: 30852548]
35. Alschuler KN, Beier ML. Intolerance of Uncertainty: Shaping an Agenda for Research on Coping with Multiple Sclerosis. Int J MS Care 2015 Jul 01;17(4):153-158 [FREE Full text] [doi: 10.7224/1537-2073.2014-044] [Medline: 26300700]
36. Lynch SG, Kroencke DC, Denney DR. The relationship between disability and depression in multiple sclerosis: the role of uncertainty, coping, and hope. Mult Scler 2001 Dec 02;7(6):411-416. [doi: 10.1177/135245850100700611] [Medline: 11795464]
37. Mullins LL, Cote MP, Fuemmeler BF, Jean VM, Beatty WW, Paul RH. Illness intrusiveness, uncertainty, and distress in individuals with multiple sclerosis. Rehabilitation Psychology 2001;46(2):139-153. [doi: 10.1037/0090-5550.46.2.139]
38. Kalb R. The emotional and psychological impact of multiple sclerosis relapses. J Neurol Sci 2007 May 15;256 Suppl 1:S29-S33. [doi: 10.1016/j.jns.2007.01.061] [Medline: 17350045]
39. LaRocca NG, Kalb RC, Foley FW, Caruso LS. Assessment of Psychosocial Outcomes. Neurorehabilitation and Neural Repair 1993 Jan 01;7(3-4):109-116. [doi: 10.1177/136140969300700305]
40. Sá MJ. Psychological aspects of multiple sclerosis. Clinical Neurology and Neurosurgery 2008 Nov;110(9):868-877. [doi: 10.1016/j.clineuro.2007.10.001]
41. Jopson NM, Moss-Morris R. The role of illness severity and illness representations in adjusting to multiple sclerosis. Journal of Psychosomatic Research 2003 Jun;54(6):503-511. [doi: 10.1016/s0022-3999(02)00455-5]
42. Isaksson A, Ahlström G. From Symptom to Diagnosis: Illness Experiences of Multiple Sclerosis Patients. Journal of Neuroscience Nursing 2006;38(4):229-237. [doi: 10.1097/01376517-200608000-00005]
43. Johnson J. On receiving the diagnosis of multiple sclerosis: managing the transition. Multiple Sclerosis 2003 Feb 1;9(1):82-88. [doi: 10.1191/1352458503ms856oa]
44. Hakim EA, Bakheit AM, Bryant TN, Roberts MW, McIntosh-Michaelis SA, Spackman AJ, et al. The social impact of multiple sclerosis - a study of 305 patients and their relatives. Disability and Rehabilitation 2009 Jul 07;22(6):288-293. [doi: 10.1080/096382800296755]
45. Sesel A, Sharpe L, Beadnall HN, Barnett MH, Szabo M, Naismith SL. The evaluation of an online mindfulness program for people with multiple sclerosis: study protocol. BMC Neurol 2019 Jun 14;19(1):129 [FREE Full text] [doi: 10.1186/s12883-019-1356-9] [Medline: 31200686]

Abbreviations

CBT: cognitive behavior therapy
CES-D: Center for Epidemiological Studies for Depression questionnaire
HRQoL: health-related quality of life
MBSR: mindfulness-based stress reduction
MS: multiple sclerosis
RCT: randomized controlled trial
STAI: State-Trait Anxiety Inventory

Edited by G Eysenbach; submitted 17.04.20; peer-reviewed by L Beatty, M Ardebil; comments to author 26.07.20; revised version received 05.08.20; accepted 26.10.20; published 02.03.21.

Please cite as:
Sesel AL, Sharpe L, Beadnall HN, Barnett MH, Szabo M, Naismith SL
Development of a Web-Based Mindfulness Program for People With Multiple Sclerosis: Qualitative Co-Design Study
J Med Internet Res 2021;23(3):e19309
URL: https://www.jmir.org/2021/3/e19309
doi:10.2196/19309
PMID:33650980

Original Paper

Evaluating a Hybrid Web-Based Training Program for Panic Disorder and Agoraphobia: Randomized Controlled Trial

Lara Ebenfeld[1], PhD; Dirk Lehr[1], Prof Dr; David Daniel Ebert[1,2], Prof Dr; Stefan Kleine Stegemann[1], MSc; Heleen Riper[3,4,5], Prof Dr; Burkhardt Funk[1], Prof Dr; Matthias Berking[6], Prof Dr

[1]Leuphana University, Lüneburg, Germany

[2]Technical University of Munich, Munich, Germany

[3]GGZ ingeest Amsterdam, Amsterdam, Netherlands

[4]Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[5]Amsterdam Public Health Research Institute, Amsterdam, Netherlands

[6]Friedrich-Alexander-University Nuremberg-Erlangen, Nuremberg-Erlangen, Germany

Corresponding Author:
Lara Ebenfeld, PhD
Leuphana University
Universitätsallee 1
Lüneburg, 21335
Germany
Phone: 49 4131 677 1708
Email: lara.ebenfeld@gmail.com

Abstract

Background: Previous studies provide evidence for the effectiveness of web-based interventions for panic disorder with and without agoraphobia. Smartphone-based technologies hold significant potential for further enhancing the accessibility and efficacy of such interventions.

Objective: This randomized controlled trial aims to evaluate the efficacy of a guided, hybrid web-based training program based on cognitive behavioral therapy for adults with symptoms of panic disorder.

Methods: Participants (N=92) with total scores in the Panic and Agoraphobia Scale ranging from 9 to 28 were recruited from the general population and allocated either to a hybrid intervention (GET.ON Panic) or to a wait-list control group. The primary outcome was the reduction in panic symptoms, as self-assessed using a web-based version of the Panic and Agoraphobia Scale.

Results: Analysis of covariance-based intention-to-treat analyses revealed a significantly stronger decrease in panic symptoms posttreatment (F=9.77; P=.002; Cohen d=0.66; 95% CI 0.24-1.08) in the intervention group than in the wait-list control group. Comparisons between groups of the follow-up measures at 3 and 6 months yielded even stronger effects (3-month follow-up: F=17.40, P<.001, Cohen d=0.89, 95% CI 0.46-1.31; 6-month follow-up: F=14.63, P<.001, Cohen d=0.81, 95% CI 0.38-1.24).

Conclusions: Hybrid web-based training programs may help reduce the symptoms of panic disorder and hence play an important role in improving health care for patients with this debilitating disorder.

Trial Registration: German Clinical Trial Register DRKS00005223; https://tinyurl.com/f4zt5ran

International Registered Report Identifier (IRRID): RR2-10.1186/1745-6215-15-427

(J Med Internet Res 2021;23(3):e20829) doi:10.2196/20829

KEYWORDS

panic disorder; agoraphobia; treatment; internet; mobile phone; randomized controlled trial

Introduction

With a 12-month prevalence of 1.8% among adults, panic disorder is one of the most common anxiety disorders [1,2]. Subthreshold cases, defined as significant panic symptoms that fail to meet full criteria, have been estimated to be just as prevalent [3,4] and have been shown to predict the development of full panic disorder as well as other mental disorders, such as generalized anxiety disorder or major depression [5]. Effective treatments for panic disorder and associated agoraphobic

symptoms include pharmacotherapy and cognitive behavioral therapy (CBT) [6-9]. Unfortunately, many individuals still lack access to evidence-based treatments because of the limited availability of clinicians or fear of stigmatization [10-12].

Technology-based psychological interventions that use the internet provide low-threshold access to evidence-based mental health care. Recent outcome studies [13-15], meta-analyses, and reviews [16-24] provide ample evidence that internet-based interventions based on cognitive behavioral therapy principles (iCBT) are effective in treating panic disorder.

Owing to their ability to bridge distances between patients and therapists, good cost-efficacy and low-threshold iCBT have great potential to facilitate access to evidence-based interventions [18,25]. However, the current dominance of desktop-based iCBT in research and health services neglects the dramatic shift in user preferences toward the use of smartphones [26]. Moreover, smartphones accompany their users wherever they go, thereby providing an excellent opportunity for ecological momentary assessment of relevant health information [27-31]. Furthermore, smartphones allow the use of ecological momentary interventions to be delivered in the real world and real time, ideally at the very moment the intervention is needed [32]. Considering the rapid growth and potential of mobile technology, surprisingly little research has been conducted to clarify the benefits of using smartphones as stand-alone or add-on interventions [33]. Available data often come from studies criticized for poor-quality interventions [34], and many interventions currently available have not been evaluated at all [35-37].

The few currently available studies provide preliminary evidence for the efficacy of smartphone-based interventions for the symptoms of anxiety disorders. For example, in a meta-analysis on the efficacy of transdiagnostic eHealth interventions that integrated mobile technologies, Firth et al [38] showed that such interventions can significantly reduce overall anxiety (Hedge g=0.45). A recent study by Christoforou et al [39] evaluated the efficacy of an app for agoraphobic symptoms in comparison with a stress reduction app. Although there was a significant pre- to posttest effect for the interventions (Panic and Agoraphobia Scale [PAS] difference −5.97; 95% CI −8.49 to −3.44), no significant differences between the interventions were observed.

Despite these promising findings, it is important to acknowledge that mobile apps also have some disadvantages with regard to usability issues. For example, elaborate writing tasks, a typical component of iCBT interventions, are difficult to complete on a small screen with a smartphone touchpad. Moreover, cellphones are typically used for short time intervals and often while performing other tasks. This is problematic, as working toward health-promoting changes often requires more sustained and focused effort [40,41]. Therefore, it can be argued that hybrid interventions that combine the advantages of both desktop and mobile technology should be superior to exclusively desktop- or mobile-based approaches. In hybrid interventions, the mobile component can be used to monitor symptoms and cue exercises in the patient's natural environment, whereas the desktop component provides text- and video-based psychoeducation and facilitates elaborate writing tasks.

Despite the obvious advantages of hybrid interventions, the literature on their efficacy is still scarce. In a transdiagnostic approach, Proudfoot et al [42] showed that the delivery of CBT using a combination of mobile app and desktop-based technology was effective in reducing symptoms of anxiety disorders (Cohen d=0.47) compared with a wait-list control (WLC) condition. Furthermore, in a study evaluating the combination of Acceptance and Commitment Therapy and a smartphone app for participants with panic disorder or social phobia, Ivanova et al [43] found no significant effect on panic symptom severity reduction. At this point, no study has been published on the efficacy of hybrid iCBT interventions for panic disorder. To fill this gap in the literature, this study aims to evaluate the efficacy of a newly developed hybrid iCBT training program for individuals with symptoms of panic disorder. Owing to the legal restrictions on remote treatment (*Fernbehandlungsgesetz*) [44], we use the term *online training program* for the intervention format instead of the term *online therapy*, which is more commonly used in the literature.

Methods

Study Design

To evaluate the efficacy of a hybrid web-based training program for panic disorder (with and without agoraphobia), we conducted a prospective, two-arm randomized trial, in which 92 participants with significant symptoms of panic disorder were randomly allocated either to the GET.ON Panic intervention group (IG) who received the training program immediately or to the WLC group who received the training program 6 months after randomization. For randomization, we used the automated computer program DatInf RandList version 1.2 (DatInf GmbH). The allocation was stratified for clinical or subclinical symptomatology as well as the presence or absence of agoraphobia in the order of incoming informed consent. To include equal numbers of participants in each group, we used block randomization (n=2 per block). The staff conducting the diagnostic interviews and observer ratings were blinded to the participants' randomization statuses. The participants could participate in the training program with a pseudonym of their own choice. Ethical approval for this study was obtained from the Ethical Committee of the University of Marburg (registration number: 2013-23 K). The study was preregistered with the German Clinical Trial Register (registration number: DRKS00005223). The study protocol was submitted for publication before randomization [45].

Participants and Recruitment

The study participants were recruited from the general population between August 2013 and October 2014. Announcements in newspapers, support groups, and social media, such as Facebook, guided interested individuals to the web-based health center website of our research group, where they could apply on the web to participate in the study. Applicants were asked to complete a web-based questionnaire assessing the following inclusion criteria: (1) experiencing mild-to-moderate panic symptoms as assessed by the PAS (score

range: 9-28) [46,47], (2) being aged ≥18 years, (3) having panic as the primary concern for seeking help, (4) having internet and smartphone access with minimum system requirements of iPhone (TM) 3GS (Apple Inc; iOS 6 and iOS 7) or a comparable Android device (Android 2.3 or newer), and (5) providing their informed consent.

The exclusion criteria were as follows: (1) receiving current psychological help for anxiety problems or being on a wait-list for psychotherapy; (2) having physical health problems that were assessed via a self-report that prevents participants from engaging in self-exposure, as recommended by the German guideline for treating people with panic disorder and agoraphobia [48]; (3) currently having posttraumatic stress disorder or psychotic or dissociative disorders assessed via self-report and clinical interview; and (4) having current suicidality, as assessed by a score above 1 on item 9 of the Beck Depression Inventory II (BDI-II) [49,50] and question A9 of the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition Axis I Disorders (SCID-I) [51]. In the event that potential participants were excluded because of suicidal ideation or intention, they were given information about further help according to an established suicide protocol. All excluded participants were contacted via email and provided with information regarding where they could obtain appropriate help.

Treatment

Participants in the treatment condition received the GET.ON Panic treatment, which is a hybrid (ie, desktop-based and smartphone-based), iCBT-based self-help intervention for treating symptoms of panic disorder [45]. Participants were advised to log on to the training platform, which was provided by the technical partner Minddistrict GmbH on a weekly basis and consecutively work through the following sessions: (1) psychoeducation, (2) interoceptive exposure, (3) in vivo exposure, (4) cognitive restructuring—introduction, (5) cognitive restructuring—extension, and (6) relapse prevention. In addition, participants were instructed on the complementary use of the GET.ON Panic app [52]. The app supported participants in (1) completing their homework assignments (eg, keeping an anxiety diary); (2) planning, conducting, and evaluating interoceptive and in vivo exposure tasks; and (3) performing relaxation exercises (Table 1).

After every session, participants received written feedback from a trained coach based on a coaching manual developed by members of our research group to ensure a standardized procedure of coaching (the manual is available on request). The guidance focused on increasing motivation and adherence throughout the training progress, rather than providing individual therapeutic advice. The average feedback took about 20-30 min. Coaches also sent reminders via a secure messaging system within the training platform if participants did not log on for 1 week. All coaches had a degree in psychology and were supervised by a licensed clinical psychologist.

Table 1. Overview of sessions.

Week	Content and homework	
	Browser	Mobile
1	Psychoeducation: • Information about panic • Defining goals of training • Setting up a reward list	• Daily diary • Registration of current panic event (event-based) • Daily summary of panic, avoidance, and mood
2	Interoceptive exposure: • Bodily symptoms during panic • Avoidance • Safety behaviors	• Respiratory interoceptive exposure exercises • Daily diary
3	In vivo exposure: • Defining an anxiety hierarchy	• In vivo exposures • Dizziness interoceptive exercises • Daily diary
4	Cognitive restructuring I: • Negative automatic thoughts • Defining anxiety project (training schedule for exposures)	• In vivo exposures • Further interoceptive exposure exercises • Daily diary
5	Cognitive restructuring II: • Reality testing of automatic negative thoughts	• In vivo exposures • Further interoceptive exposure exercises • Daily diary
6	Relapse prevention: • Early warning signs • Critical life events • Evaluation of training and aims	• Breathing and muscle relaxation exercises

Outcome Measures

Panic Symptom Severity and Self-Rating

The primary outcome was the severity of panic and agoraphobia symptoms, as self-assessed using the PAS (German version: Panik- und Agoraphobieskala) [46,47,53,54]. This scale consists of 13 items separated into the subscales of panic attacks, agoraphobic avoidance, anticipatory anxiety, daily life limitations, and health concerns. For each item, participants rated the frequency of panic symptoms during the past week on a 5-point scale. Thus, the total score ranges from 0 to 52, with scores ranging from 0 to 8 indicating *no clinically relevant symptoms*, scores ranging between 9 and 28 indicating *moderate symptoms*, and a score of 29 and above indicating a *severe level of symptoms*. Previous studies provide evidence for the efficacy of the measures, for example, Cronbach α=.86 [47] or α=.70 to .94 [55]. In this study, Cronbach α for the total score was .89.

Observer-Rated Anxiety Symptoms

The Hamilton Anxiety Scale (HAM-A) [56-58] was used as a complement for the self-administered anxiety scales. The scale contains 14 items, with a total score ranging from 0 to 30. Previous studies showed excellent interrater and test-retest reliabilities of intraclass correlation coefficients of 0.89-0.99 [57]. To examine interrater reliability in this trial, we audiotaped all the observer ratings. Around one-tenth (equivalent to 28 interviews) of these audio files were rated by experienced, blinded second raters. The interrater reliability was excellent, with an intraclass correlation coefficient of 0.99.

Agoraphobic Cognitions

The Agoraphobic Cognitions Questionnaire (ACQ) is a 14-item self-report questionnaire that measures agoraphobic cognitions. The total sum score of the ACQ ranged from 1 to 5. The ACQ has an internal reliability of Cronbach α of .80 [59-61]. In this trial, we found a Cronbach α of .84.

Body Sensations

Bodily sensations were measured using the Body Sensations Questionnaire (BSQ), a self-rating questionnaire with the total score ranging from 1 to 5 points. It has good internal reliability of Cronbach α of .87 [59-61]. In this trial, Cronbach α was .86.

Agoraphobic Avoidance

The Mobility Inventory (MI) is a questionnaire that measures agoraphobic avoidance. Participants were asked to rate common agoraphobic situations with regard to their avoidance. Each item is rated twice: once for dealing with the situation alone and once when accompanied. The total score ranged from 1 to 5 points. The internal consistencies reported in previous studies were very good, with Cronbach α of .91 (accompanied by significant others) and .94 (alone) [59,61,62]. In this study, Cronbach α values of the MI were .93 (accompanied) and .95 (alone).

Depressive Symptoms

We used the German adaption (ADS) of the Center for Epidemiologic Studies Depression Scale (CES-D) to assess depressive symptom severity. The CES-D measures 20 symptoms of depression in the previous week. The total score ranges from 0 to 60. Internal consistency has been shown to be good (Cronbach α=.89) [63,64]. In this study, Cronbach α was .87.

Diagnostic Status

The presence of panic disorder, any other anxiety disorder, or a current depressive episode was assessed using a telephone version of the SCID-I at the 6-month follow-up (6M-FU) assessment covering the period of the last 3 months by trained interviewers. Previous studies have shown excellent test-retest reliability between the 2 different formats, the telephone version and the face-to-face (f2f) version of the diagnostic interview (Cohen =0.84) [65-67]. To determine the interrater reliability of the diagnostic interviews, we used the statistics. In a previous study, moderate interrater reliability (Cohen =0.67) was found [68]. In this trial, all interviews were audiotaped, with 11.1% (18/162) of the interviews rated by an experienced, blinded second rater. Agreement between the 2 raters was moderate, with a Cohen of 0.51.

Quality of Life

Quality of life was measured using the 12-item Short-Form Health Survey (SF-12), which assesses 8 health domains: physical functioning, role limitations, pain, general health perception, vitality, mental health, emotional role, and social functioning. The SF-12 provides 2 summary scores for physical and mental health [69,70]. In this trial, Cronbach α was .79.

User Satisfaction

We assessed user satisfaction with the Client Satisfaction Questionnaire adapted to internet-based interventions (CSQ-I) [71], which is based on the German version of the Client Satisfaction Questionnaire [72,73]. Statements such as “I would recommend this training to a friend, if he or she was in need of similar help” are rated on a 4-point Likert scale (ranging from 1=*does not apply to me* to 4=*does totally apply to me*). The questionnaire contained 8 items, with a total score ranging from 8 to 32. The psychometric properties were excellent with a McDonald ω of 0.93 and ω of 0.95 [71]. In this trial, McDonald ω was 0.97.

App Usage

The mobile app contains a diary for recording and monitoring panic-related symptoms, such as panic events, degree of avoidance behavior, general anxiety, and mood level on a visual analog scale (0-10). Furthermore, the app recorded the type and number of exposure exercises performed by the participant. In addition, we used the System Usability Scale (SUS) at postassessment (T2) to assess the usability of the GET.ON Panic app [74,75]. The sum score ranges from 0 to 100, with a higher score indicating better usability.

Assessment Schedule

Participants completed a sociodemographic questionnaire, the PAS and the Suicidality item of BDI-II at screening (T0); at baseline (T1), we assessed the PAS, the SCID-I, the HAM-A, the ACQ, the BSQ, the MI, the CES-D, and the SF-12; at postassessment (T2), we assessed the PAS, the HAM-A, the ACQ, the BSQ, the MI, the CES-D, the SF-12, the CSQ-I (only

IG), and the SUS (only IG); at 3-month follow-up (3M-FU; T3), we assessed the PAS, the ACQ, the BSQ, the MI, the CES-D, and the SF-12; and at 6M-FU (T4), we assessed the PAS, the SCID-I, the HAM-A, the ACQ, the BSQ, the MI, the CES-D, and the SF-12. Diary data were continuously measured during the training period and beyond (Table 2).

Table 2. Assessment schedule.

Assessments	Screening	T1	T2	T3	T4
Sociodemographic questionnaire	X[a]	—[b]	—	—	—
Suicidality (Item 9; BDI-II[c])	X	—	—	—	—
Diagnosis (SCID-I[d], sections for anxiety disorders and current depressive episode)	—	X	—	—	X
Panic and agoraphobia severity, self-rating (PAS[e])	X	X	X	X	X
Panic and agoraphobia severity, observer-rating (HAM-A[f])	—	—	X	—	X
Agoraphobic cognitions (ACQ[g])	—	X	X	X	X
Body sensations (BSQ[h])	—	X	X	X	X
Agoraphobic avoidance (Mobility Inventory)	—	X	X	X	X
Depressive symptoms (CES-D[i])	—	X	X	X	X
Quality of life (SF-12[j])	—	X	X	X	X
User satisfaction (CSQ-I[k])	—	—	(X)[l]	—	—
Usability of smartphone app (SUS[m])	—	—	(X)	—	—

[a]Measured.

[b]Not measured.

[c]BDI- II: Beck Depression Inventory II.

[d]SCID-I: Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition Axis I Disorders.

[e]PAS: Panic and Agoraphobia Scale.

[f]HAM-A: Hamilton Anxiety Scale.

[g]ACQ: Agoraphobic Cognitions Questionnaire.

[h]BSQ: Body Sensations Questionnaire.

[i]CES-D: Center for Epidemiologic Studies Depression Scale.

[j]SF-12: 12-Item Short-Form Health Survey.

[k]CSQ-I: Client Satisfaction Questionnaire adapted to internet-based interventions.

[l]Only available for intervention group.

[m]SUS: System Usability Scale.

Statistical Analyses

To assess treatment efficacy, the GET.ON Panic group was compared with the WLC group on all outcome measures (T2, T3, and T4) using univariate analyses of covariance (ANCOVAs) with the baseline scores as covariates. On the basis of a previous meta-analysis [21], we powered the study to detect an effect size of Cohen *d*=0.6 (1− of 80%; =.05) with intention-to-treat (ITT) at T2 as our primary level of analysis. Accordingly, a sample size of 90 was required. Cohen *d* [76] was used to measure effect size. To account for covariance, we calculated Cohen *d* over the partial eta squared (η^2). To assess a clinically reliable change of panic severity (response) on an individual level, we calculated the Reliable Change Index (RCI) as proposed by Jacobson and Truax [77], coded participants as responders or deteriorators if their score on the PAS differed by 10.68 points on the PAS, and performed a Pearson chi-square test to compare the reliable change of the GET.ON Panic group with the WLC. Corresponding to the RCI, we calculated numbers needed to treat (NNT) score indicating how many participants must take part in order for GET.ON Panic to achieve one clinically relevant improvement. To assess remission rates, we calculated the percentage of people who had a diagnostic status of panic disorder according to the SCID-I interview at baseline (T0) and at the 6M-FU (T4) and performed a Pearson chi-square test to compare the diagnostic status of the GET.ON Panic group with the WLC covering a period of the last 3 months.

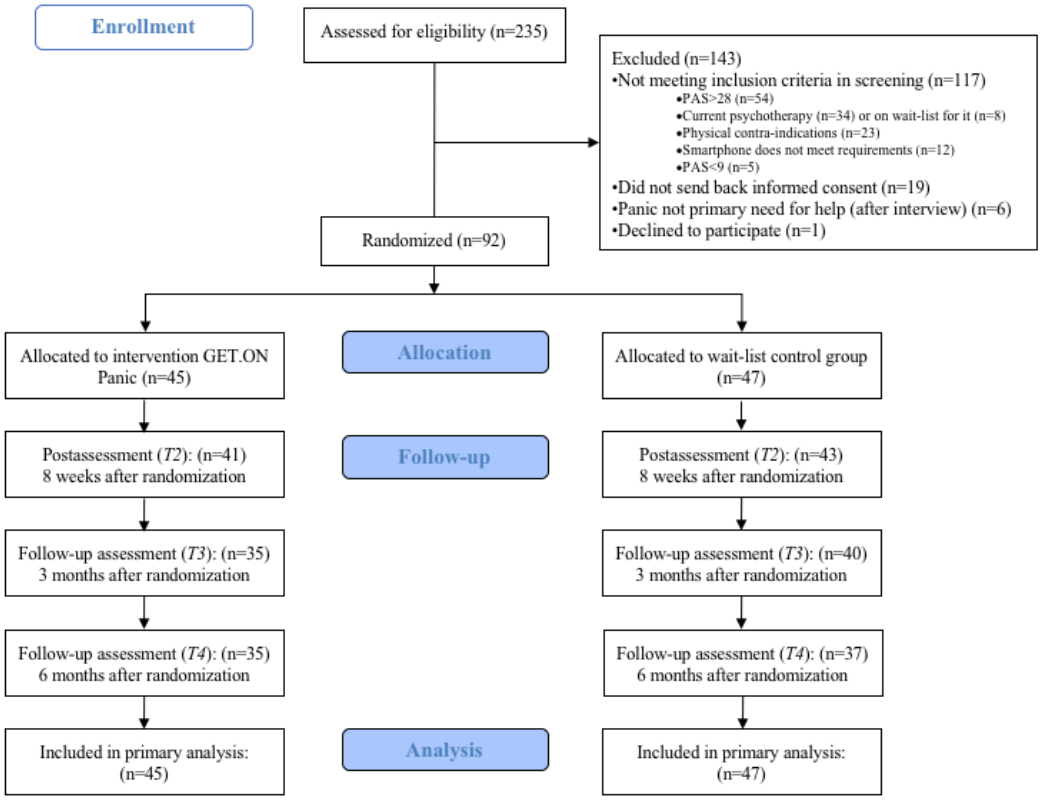
Missing data at postassessment, 3M-FU assessment, and 6M-FU assessment were performed using a Markov Chain Monte Carlo multivariate imputation algorithm (SPSS 23) with hundred estimations per missing value and all available data on outcomes, age, and gender as predictors [78].

Results

Enrollment

Over a period of 14 months, a total of 235 individuals completed the screening questionnaire. Of these, 117 did not meet the inclusion criteria or matched one or more exclusion criteria (Figure 1). Severe panic symptom severity (n=54), current psychotherapy (n=34), or physical contraindications (n=23) were the most frequent reasons for exclusion. The remaining 118 candidates were eligible to participate in the clinical interviews. Of those, 19 did not provide informed consent. After this interview, another 6 candidates were excluded because they did not have panic symptoms as their primary reason for seeking help. All excluded individuals were provided with information about applicable health care system services. The remaining 92 participants were randomly assigned to the hybrid web-based training program GET.ON Panic or the WLC condition.

Figure 1. Study flow. PAS: Panic and Agoraphobia Scale.



Baseline Characteristics

Most participants were female (51/92, 55%), White (76/92, 83%), on average aged 38 years (SD 10.42), highly educated (60/92, 65%), married or in a relationship (82/92, 89%), and currently working (51/92, 55%). On the basis of the SCID interview, the most common diagnosis was panic disorder with agoraphobia (78/92, 83%). A significant number of patients (12/92, 13%) met the criteria for panic disorder without agoraphobia. Of all participants, 26% (24/92) met the criteria for at least one additional anxiety disorder, in addition to panic disorder. A small percentage (2/92, 2%) had a current major depressive episode as a comorbid condition. Most participants (58/92, 63%) reported that they had previously undergone psychotherapeutic treatment (Multimedia Appendix 1).

Study Dropout and Compliance in Treatment

Baseline data were available for all the participants. The attrition rate was 9% (8/92) at postassessment (4/45 in the IG and 4/47 in the WLC), 18% (17/92) at the 3M-FU (10/45 in the IG and 7/47 in the WLC), and 22% (20/92) at the 6M-FU (10/45 in the IG and 10/47 in the WLC; Figure 1).

On average, the number of completed sessions in the GET.ON Panic group was 5.11 (SD 1.67). All 6 sessions were completed by 73% (33/45) of the participants, 4% (2/45) of the participants completed only session 1, 11% (5/45) of the participants dropped out after session 2, 7% (3/45) of the participants were lost after session 3, 2% (1/45) of the participants stopped the training after completing session 4, and 2% (1/45) of the participants after session 5. In total, 27% (12/45) of the participants did not complete the training. The reasons for intervention dropout were mentioned for 33% (4/12) of them: lack of time, lack of motivation, lack of personal contact with the eCoach, or surgery that interfered with completing the intervention. The resting 67% (8/12) of the participants were also study dropouts, and no reasons for stopping the intervention were known because they did not complete the postassessment.

Severity of Panic Symptoms

Preliminary analyses indicated that all necessary conditions for the intended statistical analyses were met. There was a greater decrease in self-reported panic disorder symptom severity in the intervention condition than in the WLC condition (Multimedia Appendices 2 and 3). With regard to the primary outcome, participants in the GET.ON Panic condition reported

significantly lower (baseline-controlled) panic symptom severity at posttreatment than the WLC group (F=9.77; P=.002; partial η^2=0.10; Cohen d=0.66; 95% CI 0.24-1.08). This effect became even stronger (F=17.40; P<.001; partial η^2=0.16; Cohen d=0.89; 95% CI 0.46-1.31) at the 3M-FU and remained significant (F=14.63; P<.001; partial η^2=0.14; Cohen d=0.81; 95% CI 0.38-1.24) at the long-term 6M-FU. The effect sizes ranged from medium to large.

With regard to observer-based ratings, ANCOVA showed a significant difference in anxiety symptoms between groups as measured by the HAM-A at postmeasurement (F=3.97; P=.049; partial η^2=0.04; Cohen d=0.42; 95% CI 0.01-0.84) and at the 6M-FU (F=4.86; P=.03; partial η^2=0.05; Cohen d=0.47; 95% CI 0.05-0.88) with small-to-medium effect sizes. Further analyses indicated that the findings did not significantly change when the analyses were based on the study completer instead of the ITT sample.

With regard to response, the reliable clinical changes were not significant at postmeasurement (χ^2_2 [n=92]=2.5; P=.28; improvement: GET.ON Panic: 12/45, 27% and WLC: 7/47, 15%; deterioration: GET.ON Panic: 2/45, 4% and WLC: 1/47, 2%) or at the 3M-FU (χ^2_2 [n=92]=5.3; P=.07; improvement: GET.ON Panic: 14/45, 31% and WLC: 6/47, 13%; deterioration: GET.ON Panic: 0/45, 0% and WLC: 1/47, 2%). However, the GET.ON Panic group was superior to the WLC in terms of the percentage of participants attaining reliable clinical change (RCI=±10.68) in panic symptom severity at the 6M-FU (χ^2_2 [n=92]=6.0; P=.049; improvement: GET.ON Panic: 22/45, 49% and WLC: 12/47, 26%; deterioration: GET.ON Panic: 0/45, 0% and WLC: 1/47, 2%). These reliable clinical changes correspond to NNT from baseline to posttreatment of 8.49 (95% CI 3.54 to >10^6), at the 3M-FU of NNT=5.45 (95% CI 2.87-55.78), and at the 6M-FU of NNT=4.28 (95% CI 2.35-24.07). Regarding the long-term effect, these results indicate that 4 individuals had to participate in the GET.ON Panic training program to result in one additional individual having reliable clinical improvement in panic symptom severity.

With regard to remission rates, at baseline, nearly all participants (90/92, 98%) fulfilled the diagnostic criteria for panic disorder. At the 6M-FU, 76% of the participants (70/92, 76% GET.ON Panic group: 33/45, 73%; WLC: 37/47, 79%) agreed to the telephone-administered diagnostic interview. In total, 21% (15/70) of the participants were free of a diagnosis. There was a greater reduction in diagnoses in the GET.ON Panic group (11/33, 33%) than in the WLC group (4/37, 11%; χ^2_1[n=70]=5.3; P=.02).

Additional Anxiety Measures

Comparing the GET.ON Panic with the WLC group on further self-rated anxiety measurements, we found stronger between-group effect sizes for agoraphobic cognitions (partial η^2=0.06; Cohen d=0.51; 95% CI 0.05-0.93) and bodily sensations (partial η^2=0.05; Cohen d=0.46; 95% CI 0.05-0.88) in the GET.ON Panic group than in the WLC group at posttreatment. With regard to follow-up measurements, these effects remained stable for both agoraphobic cognitions (partial η^2=0.07; Cohen d=0.55; 95% CI 0.14-0.97 after 3 months and partial η^2=0.05; Cohen d=0.46; 95% CI 0.04-0.87 after 6 months) and bodily sensations (partial η^2=0.14; Cohen d=0.79; 95% CI 0.37-1.22 after 3 months and partial η^2=0.09; Cohen d=0.66; 95% CI 0.22-1.06 after 6 months). A difference in agoraphobic avoidance between the groups could only be found when participants had to manage difficult situations when they were alone with small effect sizes (partial η^2=0.05; Cohen d=0.45; 95% CI 0.04-0.86) at posttreatment and a medium effect size (partial η^2=0.11; Cohen d=0.70; 95% CI 0.27-1.12) at the 6M-FU. ANCOVA did not reveal a significant difference between the groups regarding agoraphobic avoidance when participants had to manage difficult situations when they were in companionship with other people (Multimedia Appendices 2 and 3).

Additional Measures

At the postmeasurement as well as at the 3M-FU, the GET.ON Panic group showed no significant reduction in depressive symptoms compared with the WLC group (Multimedia Appendices 2 and 3). However, at the 6M-FU, the depressive symptoms of the GET.ON Panic group decreased significantly with a small effect size compared with the WLC group (partial η^2=0.06; Cohen d=0.49; 95% CI 0.07-0.90). The results on the quality of life scales with regard to mental health showed no reduction at postmeasurement or at the 3M-FU but a medium reduction after 6 months (partial η^2=0.11; Cohen d=0.70; 95% CI 0.28-1.12). Furthermore, no differences in symptoms regarding physical aspects of quality of life were found.

App Usage and User Satisfaction

The participants of the training group (n=45) used the mobile diary on average 25.02 times (SD 19.48; range 0-56) during the 8-week training period on average (0.45 diary entries per day per participant). The repeated analysis of variance did not reveal any changes in the diary scores over a period of 8 weeks. Furthermore, they were not related to the primary outcome. The participants performed an average of 149.80 (SD 279.34; range 0-1702) interoceptive exposure exercises and 6.63 in vivo exercises (SD 17.74; range 0-113). The mean SUS score was 71.16 (SD 18.97) at posttreatment, which indicates good usability of the GET.ON Panic app. Overall, user satisfaction with the hybrid training program was high (mean 28.10, SD 5.09). For example, 91% of the participants indicated that they would recommend the training program to a friend in need.

Discussion

Principal Findings

This study aims to evaluate the efficacy of GET.ON Panic, a guided, mobile- and web-based CBT training program for adults with significant panic disorder symptoms. The results show that individuals treated with GET.ON Panic experienced a significantly greater reduction in panic disorder symptom severity than did participants in a WLC condition with between-group effect sizes of Cohen d. The findings also show that the effects were not only stable over time but even increased

after the treatment was completed (Cohen *d*/NNT=0.66/8.49 at posttreatment vs Cohen *d*/NNT=0.89/5.45 after 3 months and Cohen *d*/NNT=0.81/4.28 after 6 months).

Comparison With Prior Work

As such, they fall well into the range of reported effect sizes in meta-analyses for internet-based interventions for panic disorder (eg, Hedge *g*=0.83 [18]; *g*=1.31 [19]; Hedge *g*=0.83 [21]; Cohen *d*=0.96 [24]). The findings also showed that one of 3 participants in the IG had attained complete remission of panic disorder at the last assessment point, whereas this was only the case in one of 10 participants in the control group. With regard to secondary outcomes, it is of note that the 6M-FU effects on depressive symptoms are larger than the average of effects reported for psychological treatment for depression (Cohen *d*=0.49 vs Hedge *g*=0.35 [79]). Finally, in this study, adherence rates and user satisfaction were slightly higher than those reported in previous studies (adherence: 73% vs 66%; satisfaction: 91% vs 86% [19]).

A potential step-up could be the use of an intervention that integrates hybrid web-based training program into f2f CBT [80]. In such blended interventions, therapists might fully exploit the potential of using the ecological momentary assessment data provided by the smartphone as well as the potential of ecological momentary interventions derived from individual case formulations and carried into the patient's life with the help of their mobile devices [81]. The adherence and usability rates in this study appear to be superior to what is currently reported for desktop-based iCBT interventions. This suggests that the integration of mobile components into iCBT should be a focus of future studies. The rapidly shifting use of mobile- instead of desktop-based devices underlines this [26,82].

The finding that depressive symptom severity was significantly reduced in the IG is important, as many individuals with panic disorder also have other mental health problems such as depression [83]. As cooccurring disorders may mutually help maintain each other [5,84], it is important that comorbid conditions are treated along with the primary disorder. The positive effects of the hybrid intervention evaluated in this study on depression are consistent with the findings from CBT that successfully treat panic disorder, which also result in a reduction of depressive symptoms [85,86].

Strengths and Limitations

To our knowledge, this is the first study to examine the efficacy of iCBT training program that makes use of mobile components in a target group of people with mild-to-moderate panic and agoraphobia symptoms. One of the main strengths of this study is its solid study design, which tests a newly developed training program within a randomized controlled trial against a WLC. In addition to self-rating outcomes, we conducted clinical interviews with regard to symptom severity and changes in diagnostic status over a period of 6 months and an observer-rated anxiety outcome to validate the outcomes based on self-ratings. Furthermore, this study has high ecological validity, as participants used their own smartphones. They were supposed to interact with their smartphones as they would normally do. This may lead to a higher acceptance of and satisfaction with the GET.ON Panic training program and foster the integration of psychological interventions into the daily lives of individuals. The overall low dropout rates in this study support this assumption.

This study has several limitations that need to be considered. First, the study results cannot be generalized to all individuals with panic disorder symptoms. Participants who took part in this trial actively participated and underwent an extended eligibility procedure before the study. Many interested individuals were excluded based on the criteria defined in the study protocol. Thus, we assume that the current participants represent a more intrinsically motivated study sample and, in addition, have a higher affinity for the internet than the average individuals with panic disorder. Therefore, the external validity of this study might be limited. Second, for future treatment development, it would have been of interest to compare the hybrid intervention with both an exclusively desktop-based and an exclusively mobile-based intervention for panic disorder. However, as such a design was beyond what we could realize in this study, it would need to be used in subsequent studies. Such studies should also compare the efficacy (and cost-effectiveness) of desktop-, mobile-based, hybrid, and blended interventions with f2f therapy for panic disorder. Third, we cannot draw any conclusions on the efficacy beyond the 6M-FU assessment. Thus, future studies should evaluate the long-term effects of hybrid iCBT interventions for panic disorder.

Conclusions

The results of this study suggest that a significant number of individuals with symptoms of panic disorder can be helped with an intervention that is comparatively easy to disseminate and that can be used anonymously, which arguably lowers an important barrier to service utilization [87]. However, the results also show that about two-thirds of the participants had not completely recovered after the intervention. Thus, interventions such as GET.ON Panic might best be used in a step-by-step care framework in which patients failing to attain recovery through an internet-based intervention subsequently receive more intense (and costly) interventions [88].

Acknowledgments

This project was funded by the European Union (project number: EFRE CCI 2007DE161PR001).

Authors' Contributions

LE, DL, DE, SK, HR, BF, and MB contributed to the study design. LE, DL, DE, and MB developed the hybrid web-based training program for treating people with panic disorder with or without agoraphobia (GET.ON PANIC). SK developed the app (GET.ON

PANIC APP), which supervised the technical aspects of the intervention. LE performed statistical analyses. DL and SK contributed to statistical analyses. LE drafted the manuscript. MB revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

DL, DE, and BF are shareholders, and LE is an employee of the transfer institute (GET.ON Institute) that has usage and exploitation rights of the GET.ON Panic training program. However, at the time of developing and evaluating the app and the hybrid web-based training program GET.ON Panic, as well as drafting the manuscript, the company was not yet incorporated.

Multimedia Appendix 1
Characteristics of the study participants (N=92) allocated to online training GET.ON intervention group (n=45) and wait-list control group (WLC) (n=47) at baseline.
[DOCX File , 40 KB - jmir_v23i3e20829_app1.docx]

Multimedia Appendix 2
Mean and SD values of all outcome variables at baseline, posttreatment, and at 3-month and 6-month follow-ups (intention-to-treat, N=92).
[DOCX File , 40 KB - jmir_v23i3e20829_app2.docx]

Multimedia Appendix 3
Differences between groups at T2, T3, and T4 (intention-to-treat, N=92).
[DOCX File , 42 KB - jmir_v23i3e20829_app3.docx]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1258 KB - jmir_v23i3e20829_app4.pdf]

References

1. Wittchen HU, Heinig I, Beesdo-Baum K. Anxiety disorders in DSM-5: an overview on changes in structure and content. Nervenarzt 2014 May 17;85(5):548-552. [doi: 10.1007/s00115-013-3986-2] [Medline: 24737036]
2. Wittchen HU, Jacobi F, Rehm J, Gustavsson A, Svensson M, Jönsson B, et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. Eur Neuropsychopharmacol 2011 Sep;21(9):655-679. [doi: 10.1016/j.euroneuro.2011.07.018] [Medline: 21896369]
3. Batelaan N, De Graaf R, Van Balkom A, Vollebergh W, Beekman A. Thresholds for health and thresholds for illness: panic disorder subthreshold panic disorder. Psychol Med 2006 Nov 01;37(2):247-256. [doi: 10.1017/s0033291706009007]
4. Kessler RC, Chiu WT, Jin R, Ruscio AM, Shear K, Walters EE. The epidemiology of panic attacks, panic disorder, and agoraphobia in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2006 Apr;63(4):415-424 [FREE Full text] [doi: 10.1001/archpsyc.63.4.415] [Medline: 16585471]
5. Kinley DJ, Walker JR, Enns MW, Sareen J. Panic attacks as a risk for later psychopathology: results from a nationally representative survey. Depress Anxiety 2011 May 11;28(5):412-419. [doi: 10.1002/da.20809] [Medline: 21400640]
6. Bandelow B, Lichte T, Rudolf S, Wiltink J, Beutel M. S3-Leitlinie: kurzfassung der empfehlungen zur behandlung von angststörungen. Berlin: Springer; 2015:978-973.
7. National Institute for Health and Care Excellence. Generalised anxiety disorder and panic disorder in adults: management. Guidance and guidelines : NICE. 2011. URL: https://www.nice.org.uk/guidance/cg113/ifp/chapter/What-treatments-should-I-be-offered-for-panic-disorder [accessed 2021-02-16]
8. Sánchez-Meca J, Rosa-Alcázar AI, Marín-Martínez F, Gómez-Conesa A. Psychological treatment of panic disorder with or without agoraphobia: a meta-analysis. Clin Psychol Rev 2010 Feb;30(1):37-50. [doi: 10.1016/j.cpr.2009.08.011] [Medline: 19775792]
9. van Balkom AJ, Bakker A, Spinhoven P, Blaauw BM, Smeenk S, Ruesink B. A meta-analysis of the treatment of panic disorder with or without agoraphobia: a comparison of psychopharmacological, cognitive-behavioral, and combination treatments. J Nerv Ment Dis 1997 Aug;185(8):510-516. [doi: 10.1097/00005053-199708000-00006] [Medline: 9284865]
10. Andrews G, Issakidis C, Carter G. Shortfall in mental health service utilisation. Br J Psychiatry 2001 Nov 02;179(5):417-425. [doi: 10.1192/bjp.179.5.417] [Medline: 11689399]
11. Coles ME, Coleman SL. Barriers to treatment seeking for anxiety disorders: initial data on the role of mental health literacy. Depress Anxiety 2010 Jan;27(1):63-71. [doi: 10.1002/da.20620] [Medline: 19960488]
12. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2011 Aug;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]

13. Allen AR, Newby JM, Mackenzie A, Smith J, Boulton M, Loughnan SA, et al. Internet cognitive-behavioural treatment for panic disorder: randomised controlled trial and evidence of effectiveness in primary care. B J Psych Open 2016 Mar;2(2):154-162 [FREE Full text] [doi: 10.1192/bjpo.bp.115.001826] [Medline: 27703768]
14. Berger T, Urech A, Krieger T, Stolz T, Schulz A, Vincent A, et al. Effects of a transdiagnostic unguided internet intervention ('velibra') for anxiety disorders in primary care: results of a randomized controlled trial. Psychol Med 2017 Jan;47(1):67-80. [doi: 10.1017/S0033291716002270] [Medline: 27655039]
15. Ciuca AM, Berger T, Crişan LG, Miclea M. Internet-based treatment for panic disorder: a three-arm randomized controlled trial comparing guided (via real-time video sessions) with unguided self-help treatment and a waitlist control. PAXPD study results. J Anxiety Disord 2018 May;56:43-55. [doi: 10.1016/j.janxdis.2018.03.009] [Medline: 29625834]
16. Andersson G. Guided internet treatment for anxiety disorders. As effective as face-to-face therapies? Stud Health Technol Inform 2012 Sep 2012:181-187. [doi: 10.3233/978-1-61499-121-2-3]
17. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
18. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
19. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
20. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]
21. Haug T, Nordgreen T, Öst LG, Havik OE. Self-help treatment of anxiety disorders: a meta-analysis and meta-regression of effects and potential moderators. Clin Psychol Rev 2012 Jul;32(5):425-445. [doi: 10.1016/j.cpr.2012.04.002] [Medline: 22681915]
22. Mewton L, Smith J, Rossouw P, Andrews G. Current perspectives on internet delivered cognitive behavioral therapy for adults with anxiety and related disorders. Psychol Res and Beh Manage 2014 Jan:37. [doi: 10.2147/prbm.s40879]
23. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2015 Mar 05(3):CD011565. [doi: 10.1002/14651858.CD011565] [Medline: 25742186]
24. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
25. Hedman E, Ljótsson B, Rück C, Bergström J, Andersson G, Kaldo V, et al. Effectiveness of internet-based cognitive behaviour therapy for panic disorder in routine psychiatric care. Acta Psychiatr Scand 2013 Dec 14;128(6):457-467. [doi: 10.1111/acps.12079] [Medline: 23406572]
26. ICT facts and figures. International Telecommunication Union. 2017. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf [accessed 2021-02-08]
27. Alpers GW. Ambulatory assessment in panic disorder and specific phobia. Psychol Assess 2009 Dec;21(4):476-485. [doi: 10.1037/a0017489] [Medline: 19947782]
28. Helbig S, Lang T, Swendsen J, Hoyer J, Wittchen H. Implementierung, akzeptanz und informationsgehalt eines ecological momentary assessment (EMA)-ansatzes bei patienten mit panikstörung und agoraphobie. Zeitschrift für Klinische Psychologie und Psychotherapie 2009 Apr;38(2):108-117. [doi: 10.1026/1616-3443.38.2.108]
29. Moskowitz DS, Russell JJ, Sadikaj G, Sutton R. Measuring people intensively. Cana Psychol 2009 Aug;50(3):131-140. [doi: 10.1037/a0016625]
30. Myin-Germeys I, Oorschot M, Collip D, Lataster J, Delespaul P, van Os J. Experience sampling research in psychopathology: opening the black box of daily life. Psychol Med 2009 Feb 12;39(9):1533-1547. [doi: 10.1017/s0033291708004947]
31. Trull TJ, Ebner-Priemer UW. Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. Psychol Assess 2009 Dec;21(4):457-462 [FREE Full text] [doi: 10.1037/a0017653] [Medline: 19947780]
32. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
33. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
34. Van Singer M, Chatton A, Khazaal Y. Quality of smartphone apps related to panic disorder. Front Psychiatry 2015 Jul 14;6:96 [FREE Full text] [doi: 10.3389/fpsyt.2015.00096] [Medline: 26236242]

35. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
36. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
37. Sucala M, Cuijpers P, Muench F, Cardo R, Soflau R, Dobrean A, et al. Anxiety: there is an app for that. A systematic review of anxiety apps. Depress Anxiety 2017 Jun;34(6):518-525. [doi: 10.1002/da.22654] [Medline: 28504859]
38. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
39. Christoforou M, Sáez Fonseca JA, Tsakanikos E. Two novel cognitive behavioral therapy-based mobile apps for agoraphobia: randomized controlled trial. J Med Internet Res 2017 Nov 24;19(11):e398 [FREE Full text] [doi: 10.2196/jmir.7747] [Medline: 29175809]
40. Middleton KR, Anton SD, Perri MG. Long-term adherence to health behavior change. Am J Lifestyle Med 2013 Jun 14;7(6):395-404 [FREE Full text] [doi: 10.1177/1559827613488867] [Medline: 27547170]
41. Min JA, Lee CU, Lee C. Mental health promotion and illness prevention: a challenge for psychiatrists. Psychiatry Investig 2013 Dec;10(4):307-316 [FREE Full text] [doi: 10.4306/pi.2013.10.4.307] [Medline: 24474978]
42. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Nov 18;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
43. Ivanova E, Lindner P, Ly KH, Dahlin M, Vernmark K, Andersson G, et al. Guided and unguided acceptance and commitment therapy for social anxiety disorder and/or panic disorder provided via the internet and a smartphone application: a randomized controlled trial. J Anxiety Disord 2016 Dec;44:27-35. [doi: 10.1016/j.janxdis.2016.09.012] [Medline: 27721123]
44. Bundesärztekammer. Beschlüsse des 114. Deutschen Ärztetages in Kiel, Stand. 2011. URL: https://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/pdf-Ordner/MBO/MBO_Synopse.pdf [accessed 2011-08-29]
45. Ebenfeld L, Stegemann SK, Lehr D, Ebert DD, Jazaieri H, van Ballegooijen W, et al. Efficacy of a hybrid online training for panic symptoms and agoraphobia: study protocol for a randomized controlled trial. Trials 2014 Nov 04;15(1):427 [FREE Full text] [doi: 10.1186/1745-6215-15-427] [Medline: 25370504]
46. Bandelow B. Assessing the efficacy of treatments for panic disorder and agoraphobia. II. The panic and agoraphobia scale. Int Clin Psychopharmacol 1995 Jun;10(2):73-81. [doi: 10.1097/00004850-199506000-00003] [Medline: 7673659]
47. Bandelow B. Panic and Agoraphobia Scale (PAS). 1997. URL: https://www.testzentrale.de/shop/panik-und-agoraphobie-skala.html [accessed 2021-02-16]
48. Schneider S, Margraf J. Agoraphobia and panic disorder. URL: https://www.hogrefe.com/de/shop/agoraphobie-und-panikstoerung-76509.html [accessed 2021-02-16]
49. Beck AT, Steer RA, Brown GK. Beck depression inventory-second edition. BDI-II. URL: https://www.nctsn.org/measures/beck-depression-inventory-second-edition [accessed 2021-02-16]
50. Hautzinger M, Keller F, Kühner C. Beck Depression Inventory Revision. BDI-II. 2006. URL: https://www.testzentrale.de/shop/beck-depressions-inventar.html [accessed 2021-02-16]
51. Wittchen HU, Zaudig M, Fydrich T. SKID: structured clinical interview for DSM-IV; Axis I and II. In: Zeitschrift für Klinische Psychologie und Psychotherapie. Göttingen: Hogrefe; Jan 1999:68-70.
52. Kleine Stegemann S, Ebenfeld L, Lehr D, Berking M, Funk B. Development of a mobile application for people with panic disorder as augmentation for an internet-based intervention. In: Proceedings of the Federated Conference on Computer Science and Information Systems. 2013 Presented at: 2013 Federated Conference on Computer Science and Information Systems; Sept 8-11, 2013; Krakow, Poland p. 1331-1337.
53. Bandelow B, Broocks A, Pekrun G, George A, Meyer T, Pralle L, et al. Pharmacopsychiatry 2000 Sep;33(5):174-181. [doi: 10.1055/s-2000-12982] [Medline: 11071019]
54. Bandelow B, Hajak G, Holzrichter S, Kunert HJ, Rüther E. Assessing the efficacy of treatments for panic disorder and agoraphobia. I. Methodological problems. Int Clin Psychopharmacol 1995 Jun;10(2):83-93. [doi: 10.1097/00004850-199506000-00004] [Medline: 7673660]
55. Wölk J, Sütterlin S, Koch S, Vögele C, Schulz SM. Enhanced cardiac perception predicts impaired performance in the Iowa Gambling Task in patients with panic disorder. Brain Behav 2014 Mar 22;4(2):238-246 [FREE Full text] [doi: 10.1002/brb3.206] [Medline: 24683516]
56. Hamilton M. The assessment of anxiety states by rating. Br J Med Psychol 1959;32(1):50-55. [doi: 10.1111/j.2044-8341.1959.tb00467.x] [Medline: 13638508]
57. Shear MK, Vander Bilt J, Rucci P, Endicott J, Lydiard B, Otto MW, et al. Reliability and validity of a structured interview guide for the Hamilton Anxiety Rating Scale (SIGH-A). Depress Anxiety 2001;13(4):166-178. [Medline: 11413563]

58. Weyer G. The Hamilton-Anxiety-Scale. In: Collegium Internationale Psychiatriae Scalarum; Internationale Skalen für Psychiatrie. 6th ed. Göttingen: Beltz-Test-GmbH; 2015.
59. Chambless DL, Caputo GC, Bright P, Gallagher R. Assessment of fear of fear in agoraphobics: The Body Sensations Questionnaire and the Agoraphobic Cognitions Questionnaire. J Consult and Clin Psychol 1984 Dec;52(6):1090-1097. [doi: 10.1037/0022-006x.52.6.1090]
60. Craske MG, Rachman SJ, Tallman K. Mobility, cognitions, and panic. J Psychopathol Behav Assess 1986 Sep;8(3):199-210. [doi: 10.1007/bf00959832]
61. Ehlers A, Margraf J, Chambless D. Body-related fears, cognitions, and avoidance questionnaires. AKV. URL: https://www.testzentrale.de/shop/fragebogen-zu-koerperbezogenen-aengsten-kognitionen-und-vermeidung.html [accessed 2021-02-16]
62. Chambless DL, Caputo G, Jasin SE, Gracely EJ, Williams C. The mobility inventory for agoraphobia. Beh Res and Therapy 1985;23(1):35-44. [doi: 10.1016/0005-7967(85)90140-8]
63. Hautzinger M, Bailer M. General depression scale. ADS. URL: https://www.testzentrale.de/shop/allgemeine-depressionsskala.html [accessed 2021-02-16]
64. Radloff LS. The CES-D Scale. Applied Psychological Measurement 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]
65. Crippa JAS, Osório FDL, Del-Ben CM, Filho AS, da Silva Freitas MC, Loureiro SR. Comparability between telephone and face-to-face structured clinical interview for DSM-IV in assessing social anxiety disorder. Perspect Psychiatr Care 2008 Oct;44(4):241-247. [doi: 10.1111/j.1744-6163.2008.00183.x] [Medline: 18826462]
66. Irvine A, Drew P, Sainsbury R. 'Am I not answering your questions properly?' Clarification, adequacy and responsiveness in semi-structured telephone and face-to-face interviews. Qual Res 2012 Apr 05;13(1):87-106. [doi: 10.1177/1468794112439086]
67. Rohde P, Lewinsohn PM, Seeley JR. Comparability of telephone and face-to-face interviews in assessing axis I and II disorders. Am J Psychiatry 1997 Nov;154(11):1593-1598. [doi: 10.1176/ajp.154.11.1593] [Medline: 9356570]
68. Lobbestael J, Leurgans M, Arntz A. Inter-rater reliability of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I) and Axis II Disorders (SCID II). Clin Psychol Psychother 2011 Jan 12;18(1):75-79. [doi: 10.1002/cpp.693] [Medline: 20309842]
69. Gandek B, Ware JE, Aaronson NK, Apolone G, Bjorner JB, Brazier JE, et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries. J Clin Epidemio 1998 Nov;51(11):1171-1178. [doi: 10.1016/s0895-4356(98)00109-7]
70. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
71. Boß L, Lehr D, Reis D, Vis C, Riper H, Berking M, et al. Reliability and validity of assessing user satisfaction with web-based health interventions. J Med Internet Res 2016 Aug 31;18(8):e234 [FREE Full text] [doi: 10.2196/jmir.5952] [Medline: 27582341]
72. Attkisson C, Zwick R. The client satisfaction questionnaire. Evaluation and Program Planning 1982 Jan;5(3):233-237. [doi: 10.1016/0149-7189(82)90074-X]
73. Schmidt J, Lamprecht F, Wittmann WW. Satisfaction with inpatient management. Development of a questionnaire and initial validity studies. Psychother Psychosom Med Psychol 1989 Jul;39(7):248-255. [Medline: 2762479]
74. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an Adjective Rating Scale. J Usability Stud. 2009 May. URL: http://uxpajournal.org/wp-content/uploads/sites/8/pdf/JUS_Bangor_May2009.pdf [accessed 2021-02-08]
75. Brooke J. SUS: a 'Quick and Dirty' usability scale. London, UK: Taylor & Francis; 1996.
76. Cohen J. Statistical power analysis for the behavioral sciences. 2nd edition. London, UK: Taylor & Francis; 1988.
77. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult ng and Clin Psychol 1991;59(1):12-19. [doi: 10.1037/0022-006x.59.1.12]
78. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002 Jun;7(2):147-177. [Medline: 12090408]
79. Cuijpers P, Koole SL, van Dijke A, Roca M, Li J, Reynolds CF. Psychotherapy for subclinical depression: meta-analysis. Br J Psychiatry 2014 Oct 02;205(4):268-274 [FREE Full text] [doi: 10.1192/bjp.bp.113.138784] [Medline: 25274315]
80. Marks IM, Kenwright M, McDonough M, Whittaker M, Mataix-Cols D. Saving clinicians' time by delegating routine aspects of therapy to a computer: a randomized controlled trial in phobia/panic disorder. Psychol Med 2004 Jan 14;34(1):9-17. [doi: 10.1017/s003329170300878x] [Medline: 14971623]
81. Bruinsma A, Kampman M, Exterkate CC, Hendriks GJ. An exploratory study of 'blended' cognitive behavioural therapy (CBT) for patients with a panic disorder: results and patients' experiences. Tijdschr Psychiatr 2016;58(5):361-370 [FREE Full text] [Medline: 27213635]
82. mHealth: new horizons for health through mobile technologies. World Health Organization. 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2021-02-08]

83. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
84. Roy-Byrne PP, Stang P, Wittchen H, Ustun B, Walters EE, Kessler RC. Lifetime panic-depression comorbidity in the National Comorbidity Survey. Association with symptoms, impairment, course and help-seeking. Br J Psychiatry 2000 Mar 02;176(3):229-235. [doi: 10.1192/bjp.176.3.229] [Medline: 10755069]
85. Walderhaug EP, Gjestad R, Egeland J, Havik OE, Nordgreen T. Relationships between depressive symptoms and panic disorder symptoms during guided internet-delivered cognitive behavior therapy for panic disorder. Nord J Psychiatry 2019 Oct 02;73(7):417-424. [doi: 10.1080/08039488.2019.1646803] [Medline: 31373520]
86. Allen LB, White KS, Barlow DH, Shear MK, Gorman JM, Woods SW. Cognitive-Behavior Therapy (CBT) for panic disorder: relationship of anxiety and depression comorbidity with treatment outcome. J Psychopathol Behav Assess 2010 Jun 24;32(2):185-192 [FREE Full text] [doi: 10.1007/s10862-009-9151-3] [Medline: 20421906]
87. Ebert DD, Van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet- and mobile-based psychological interventions: applications, efficacy, and potential for improving mental health. European Psychologist 2018 May;23(2):167-187. [doi: 10.1027/1016-9040/a000318]
88. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. Br J Psychiatry 2005 Jan;186:11-17. [doi: 10.1192/bjp.186.1.11] [Medline: 15630118]

Abbreviations

3M-FU: 3-month follow-up
6M-FU: 6-month follow-up
ACQ: Agoraphobic Cognitions Questionnaire
ANCOVA: analysis of covariance
BDI-II: Beck Depression Inventory II
BSQ: Body Sensations Questionnaire
CBT: cognitive behavioral therapy
CES-D: Center for Epidemiologic Studies Depression Scale
CSQ-I: Client Satisfaction Questionnaire adapted to internet-based interventions
f2f: face-to-face
HAM-A: Hamilton Anxiety Scale
iCBT: internet-based interventions based on cognitive behavioral therapy principles
IG: intervention group
ITT: intention-to-treat
MI: mobility inventory
NNT: numbers needed to treat
PAS: Panic and Agoraphobia Scale
RCI: Reliable Change Index
SCID-I: Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition Axis I Disorders
SF-12: 12-Item Short-Form Health Survey
SUS: System Usability Scale
WLC: wait-list control

Edited by G Eysenbach; submitted 29.05.20; peer-reviewed by M Feijt, K Mathiasen; comments to author 29.06.20; revised version received 06.10.20; accepted 17.01.21; published 04.03.21.

Please cite as:
Ebenfeld L, Lehr D, Ebert DD, Kleine Stegemann S, Riper H, Funk B, Berking M
Evaluating a Hybrid Web-Based Training Program for Panic Disorder and Agoraphobia: Randomized Controlled Trial
J Med Internet Res 2021;23(3):e20829
URL: https://www.jmir.org/2021/3/e20829
doi:10.2196/20829
PMID:33661121

Original Paper

Association of Spontaneous and Induced Self-Affirmation With Smoking Cessation in Users of a Mobile App: Randomized Controlled Trial

Elizabeth L Seaman[1], MHS, PhD; Cendrine D Robinson[2], MPH, PhD; David Crane[3], PhD; Jennifer M Taber[4], PhD; Rebecca A Ferrer[2], PhD; Peter R Harris[5], PhD; William M P Klein[2], PhD

[1]CDC Foundation, Atlanta, GA, United States

[2]Behavioral Research Program (BRP), Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD, United States

[3]Smoke Free, 23 Ltd, London, United Kingdom

[4]Department of Psychological Sciences, Kent State University, Kent, OH, United States

[5]School of Psychology, University of Sussex, Falmer, Brighton, United Kingdom

Corresponding Author:
Elizabeth L Seaman, MHS, PhD
CDC Foundation
600 Peachtree Street NE, Suite 1000
Atlanta, GA, 30308
United States
Phone: 1 4438524139
Email: eseaman@cdcfoundation.org

Abstract

Background: Most smokers attempt to stop using cigarettes numerous times before successfully quitting. Cigarette cravings may undermine perceived competence to quit and thus constitute psychological threats to the individual's self-concept. Self-affirmation may promote smoking cessation by offsetting these threats.

Objective: This study examines whether self-affirmation is associated with smoking cessation in the context of a cessation app. Two types of self-affirmation are examined: tendency to spontaneously self-affirm, and self-affirmation inductions added to a publicly available smoking cessation app (Smoke-Free Quit Smoking Now). In addition, this study explores whether optimism and emotional states (happiness, anger, anxiousness, hopefulness, sadness) predict smoking cessation.

Methods: All users who met the inclusion criteria, provided consent to participate, and completed a baseline assessment, including all individual difference measures, were randomized to 1 of 4 conditions. Half of the participants were randomly assigned to complete a self-affirmation induction upon study entry. Orthogonally, half of the participants were randomly assigned to receive self-affirming text notifications during their quit attempt or to receive conventional notifications. The induction and the text notifications were fully automated, and all data were collected through self-assessments in the app. Self-reported smoking cessation was assessed 1 month and 3 months following study entry.

Results: The study enrolled 7899 participants; 647 completed the 1-month follow-up. Using an intent-to-treat analysis at the 1-month follow-up, 7.2% (569/7899) of participants self-reported not smoking in the previous week and 6.4% (503/7899) self-reported not smoking in the previous month. Greater tendency to spontaneously self-affirm predicted a greater likelihood of cessation (P<.001) at 1 month after controlling for smoking-related variables. Neither self-affirmation induction influenced cessation. In addition, spontaneous self-affirmation did not moderate the relationship between self-affirmation inductions and cessation. Greater baseline sadness was associated with a lower likelihood of reporting successful cessation. Optimism predicted past-week cessation at the 1-month follow-up, and both happiness and anger predicted past-month cessation at the 1-month follow-up; however, none of these potential predictors moderated the relationship between self-affirmation conditions and successful cessation.

Conclusions: Spontaneous self-affirmation may be an important psychological resource for managing threats to self-concept during the smoking cessation process. Sadness may hinder quit attempts. Future research can explicate how spontaneous versus induced self-affirmation can promote smoking cessation and examine boundary conditions for the effectiveness of disseminated self-affirmation interventions.

Trial Registration: ISRCTN Registry 56646695; https://www.isrctn.com/ISRCTN56646695

(J Med Internet Res 2021;23(3):e18433) doi:10.2196/18433

KEYWORDS

smoking cessation; smartphone; mHealth; sadness; self-affirmation; spontaneous self-affirmation; apps; mobile phone

Introduction

Background

Tobacco use remains a leading cause of preventable death and disease globally, contributing to over 7.1 million deaths annually [1]. Each year, approximately 343,000 people in the United States die from cancer related to tobacco use [2]. Many adults are motivated to quit smoking cigarettes; however, most attempts to quit are unsuccessful [3,4]. Clinical practice guidelines emphasize combining pharmacological treatments with behavioral interventions [5]. There are several empirically supported behavioral treatments for smoking cessation [6]. However, the high rate of unsuccessful quit attempts [7] suggests that there is a need for supplementary and easily disseminable behavioral interventions.

Mobile Health and Smoking Cessation

There is a growing body of literature supporting behavioral smoking cessation treatments delivered via mobile health platforms, including smartphones [8-10]. As smartphone access in the United States continues to rise, more individuals will have access to behavioral interventions delivered on smartphones. About 81% of US adults own a smartphone [11], and 72% of adult internet users have searched for health-related information on the web [12]. One international systematic review established that web-based health information seeking is common in many different countries and found that web-based health information seeking can improve patient-physician relationships [13].

Smartphone apps for smoking cessation can include a variety of theory-based intervention components that promote cessation, such as techniques to facilitate coping with craving and behavioral strategies for removing smoking-related stimuli from a smoker's house [14]. Smoke Free-Quit Smoking Now is one such mobile app for smoking cessation that includes the behavior change techniques of supporting users to take on the identity of a nonsmoker, rewarding cessation, and changing routines [15].

Self-Affirmation Theory and Applications to Smoking Cessation

Quitting smoking is a difficult endeavor, and most smokers attempt cessation many times before they successfully quit [16]. The process of attempting to quit smoking may in itself be psychologically threatening, as smokers may interpret cravings and temporary relapse during the process as indicators of lack of competence for quitting, constituting a threat to self-concept that results in negative affect [17]. When self-competence is threatened, it may undermine the cessation process by reducing motivation to quit or cessation self-efficacy. Many smoking cessation interventions bolster perceived competence to quit [18]. However, interventions to protect the self-concept are less common and may bolster the effectiveness of existing cessation interventions.

One such intervention approach is based on self-affirmation theory. According to this theory, people are highly motivated to see themselves as having *self-integrity*, which is marked by a sense of moral adequacy and competence [19]. Thus, when they experience threats to these attributes, they may respond defensively in an attempt to protect and bolster their self-integrity [20,21]. Health behavior change interventions often contain such threats because they suggest that one is volitionally engaged in a behavior that is harmful or irrational [22,23]. Defensiveness in the face of threats to self-integrity has been observed among smokers [24,25]. For example, smokers may respond to threatening cessation messages by impugning their content [26]. Even smoking cessation materials that are not explicitly threatening or loss-framed may be perceived as threatening by smokers attempting to quit or former smokers struggling with relapse. Self-affirmation theory suggests that to the extent that people can sustain views of themselves as morally adequate and competent, they will be more open to specific threats to the self. For example, smokers who are reassured about their self-integrity may be able to better face the challenges of cessation [27]. Accordingly, much research shows that when people have an opportunity to reflect on, for example, their cherished values before being exposed to threatening health information, such as a graphic warning label [24,28] or personal disease risk [29], they are more receptive to that information and may be more likely to engage in risk reduction behavior (meta-analyses [30-32]). We thus hypothesized that self-affirmation could offset the potential threats associated with quitting and, in turn, promote successful cessation.

Although evidence suggests that self-affirmation inductions can improve engagement with and efficacy of health behavior intentions, evidence is mixed for studies specifically targeting smokers. Some studies have found benefits of self-affirmation [28,33-38], including less defensiveness toward graphic warnings [28,36]. Moreover, when combined with other intervention strategies, such as motivational interviewing and cessation programs, self-affirmed individuals reduced cigarette consumption [37]. However, other studies have not found beneficial effects of self-affirmation on smoking-related outcomes for daily smokers [35,39,40]. Thus, additional research is needed to determine the effectiveness of induced self-affirmations among smokers.

In addition to research on self-affirmation inductions, some people are more likely than others to naturally or spontaneously engage in self-affirmation when feeling threatened or anxious [34,41-43]. Spontaneous self-affirmation may serve as a resource to facilitate smoking cessation because of its potential to offset

cessation-related psychological threats in ways similar to induced self-affirmation. Indeed, the tendency to report spontaneous self-affirmation has been associated with greater acceptance of threatening health information [41,44], greater health information seeking [45], and other positive health-related outcomes, including higher perceived quality of care and increased likelihood of asking questions in a medical appointment [45-47].

There is some evidence that spontaneous self-affirmation may be beneficial for smokers. In one cross-sectional study of U.S. adult smokers, spontaneous self-affirmation moderated the relationship between living in a state with smoke-free policies, which may constitute a threatening environment for smokers, and quit intentions [48]. In this study, we examined whether spontaneous self-affirmation was associated with quit outcomes among a global cohort of smokers enrolled in a UK-based smoking app. In addition, we examined whether the tendency to spontaneously self-affirm moderated the relationship between self-affirmation inductions and successful cessation.

In addition to the tendency to spontaneously self-affirm, other psychological states and individual differences may serve as resources to bolster smoking cessation, either by interacting with self-affirmation or on their own. In this study, we examined optimism and sadness. Optimism refers to a general tendency to expect positive future events [49]. As optimism is a psychological resource that can bolster goal pursuit [50,51], people with higher optimism may have greater success at smoking cessation. Spontaneous self-affirmation and optimism are distinct psychological processes [41]; however, they may have similar associations with health outcomes, extending to smoking cessation [46]. Currently experienced emotions may also influence smoking cessation; such emotions may trigger action tendencies that facilitate predictable patterns of behavior [52-54]. Sadness, in particular, may be a hindrance to quitting smoking and predicting relapse during the smoking cessation process [55]. Sadness is associated with reward-seeking tendencies to mitigate loss [56], which can result in increased hedonically pleasing, but often unhealthy, appetitive behavior [57], including smoking [55]. Thus, when current or former smokers feel sad, they may turn to cigarettes in an attempt to improve their mood. In addition to influencing cessation success, emotion may influence the experience of relapse during the smoking cessation process.

Self-Affirmation and Mobile Health: Creating Scalable Health Behavior Interventions

Health behavior interventions that are mobile or remotely delivered are easily implemented and widely disseminated, and integrating self-affirmation content could enhance their efficacy. The scalability of self-affirmation interventions has been demonstrated in other domains in which threat impedes adaptive outcomes (eg, education; [58]) but has rarely been examined in health contexts. Indeed, self-affirmation opportunities are disseminable, given that affirmation exercises require little time and effort but can have lasting effects [58,59]. Enduring effects from such a low-burden intervention are hypothesized to work through recursive processes [32]. That is, it is not necessarily the affirmation itself that continues to influence behavior over time. Rather, affirmation attenuates threats to self-competence that might arise from cravings and temporary relapse, which might otherwise impede motivation to quit smoking, and then allows individuals to capitalize on existing resources to facilitate behavior change [58,60]. A recent meta-analysis suggested that self-affirmation is more likely to facilitate change when psychological threat impedes behavior change and when resources are present to support such change [32]. Thus, implementing brief affirmations into an existing, scalable cessation intervention in which psychological threat may impede cessation may bolster the effectiveness of the intervention by maximizing the likelihood that individuals will benefit from the behavior change resources provided in the intervention.

Previous Work Informing This Study

This study was designed as a follow-up to a previous study that provided initial evidence that incorporating a self-affirmation component into a standard text message–based smoking cessation intervention was a feasible, low-cost, and potentially efficacious way of bolstering the content of that intervention [61]. This previous study used a 2 (baseline affirmation: present vs absent) x 2 (integrated affirmation texts: present vs absent) factorial design [61]. In that study, 1261 participants met the eligibility criteria and initiated the program, 687 participants remained enrolled throughout the 42-day intervention, and 81 participants reported their smoking status at the end of the 42-day intervention [61]. Although there were no significant effects of affirmations on cessation when examining participants who remained enrolled in the study (n=687), affirmations did facilitate cessation when only participants who reported their smoking status at the 42-day follow-up were included in analyses (n=81, 6.4% of eligible baseline respondents) [61]. Intent-to-treat analyses of the 1261 participants who initiated the program indicated a 5.6% cessation prevalence at the 42-day follow-up [61]. This study builds on this former work by (1) providing a test of replication and (2) examining the role of individual differences in spontaneous self-affirmation, optimism, and affect.

This Study and Hypotheses

This study was intended to replicate our team's earlier study (Taber et al [61]) in a different setting. In this study, we tested whether self-affirmation was associated with better cessation outcomes in the context of a smoking cessation app. This study had 2 primary aims: to assess the effect of induced self-affirmation conditions on smoking cessation outcomes (aim 1) and to assess the associations of spontaneous self-affirmation with smoking cessation outcomes (aim 2). We hypothesized that 2 types of self-affirmation opportunities—a baseline kindness quiz and self-affirming push notifications in the subsequent months—would promote cessation. We also hypothesized that individuals with a tendency to spontaneously self-affirm at baseline would be more likely to successfully quit smoking. In the absence of relevant findings on which to base hypotheses, we tested whether induced self-affirmation conditions were more or less effective for people higher versus lower in spontaneous self-affirmation [41]. Finally, an exploratory aim (aim 3) was to assess baseline optimism and baseline affective states (happiness, anger, anxiousness,

hopefulness, sadness) as potential predictors and potential moderators of the relationship between affirmation conditions and cessation outcomes.

Methods

Smoke Free App

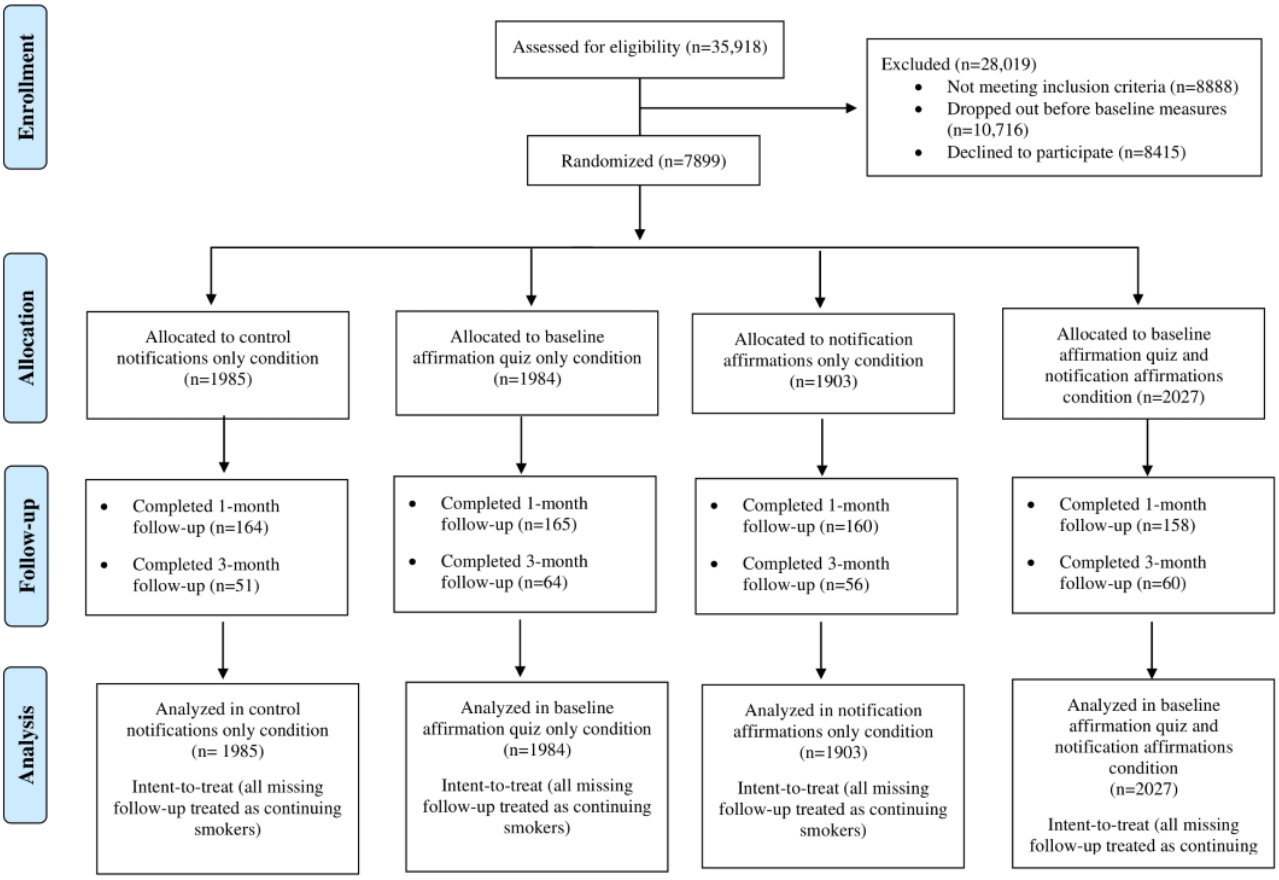
Smoke Free-Quit Smoking Now is a UK-based app designed for iOS and Android devices [15] that attracts users from across the globe. The app averages 3000 new downloads a day and allows users to set a quit date and track their craving [15]. The app offers 4 methods by which users can monitor their progress: (1) a calendar showing the total amount of time since they stopped smoking, (2) a calculator that shows the amount of money saved by not buying cigarettes and the number of cigarettes not smoked, (3) virtual badges users can earn for milestones, such as 50 hours smoke-free, and (4) health progress indicators to monitor improvements since cessation [15].

Participants, Recruitment, and Eligibility

A randomly selected proportion of users who downloaded the app during the study period (initially 10% and then increased to 30% to achieve recruitment goals) were shown a consent form and invited to participate in this study. In the informed consent form, participants were told that they could opt out of the study at any point by contacting the study investigator. The study employed a 2 (baseline self-affirmation induction: present vs absent) x 2 (notifications: self-affirming texts vs control texts) double-blind randomized controlled trial (RCT) in which self-affirmation opportunities were added to an existing smoking cessation app. Those who consented were randomly assigned to 1 of the 4 conditions and completed a baseline assessment. The initial recruitment goal was 5000 participants to have 500 completing the 1-month follow-up survey after accounting for 90% attrition, similar to previous studies [61,62]. A sample size of 500 at the 1-month follow-up was calculated to be able to detect a small effect size (F=.15), with high (.90) power using an analysis of variance (ANOVA) with 4 groups (calculated with G*Power). Participants were asked to complete 2 follow-up assessments 1 month and 3 months after they downloaded the app to assess cessation behavior and smoking status. All data were collected through the smartphone apps—users were notified about follow-up surveys in the app with one push notification and a red dot added to the app icon to indicate user action was requested. The link to complete the survey remained in the *Settings* section of the app until the participant responded.

Once participants completed the baseline assessment, their eligibility was determined. Participants were not included in the study if they were under 18 years or over 98 years of age, selected a quit date more than 14 days in the future or more than 1 day in the past, paid for additional app features (Pro users), or did not complete the baseline assessment. In a divergence from the previous study [61], potential participants in this study were required to have listed a quit date after the day they downloaded the app; this ensured that participants randomized to the baseline affirmation condition would take the baseline affirmation quiz before attempting cessation. In addition, during data collection, a glitch occurred in which the same identifier was assigned to multiple participants; all users affected by this glitch were excluded from the study and are indicated in Figure 1 under the designation of *not meeting inclusion criteria*. App users who were not eligible for the study could still use the app. All participants were entered in a lottery—noncontingent on completion of surveys—for a US $100 Amazon gift card. This study was approved by the Institutional Review Board of the National Cancer Institute.

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) diagram.



In total, 7899 participants met all inclusion criteria, were enrolled in the study, and provided survey responses. The country of participant residence was not assessed at the individual level; however, aggregate information about the geographic location of participants was available; most participants were from the United Kingdom, closely followed by the United States (Multimedia Appendix 1). Overall, the mean age of participants was 30.5 years (SD 8.7). The majority of participants were male (4790/7899, 60.6%) and did not use any cessation aids other than the Smoke Free app at baseline (6178/7899, 78.2%). Table 1 shows the demographic characteristics of participants at baseline. Significance assessed in Table 1 used a Bonferroni-corrected α level (.05 and 16 comparisons, so the adjusted α level is $P<.003125$).

Table 1. Baseline characteristics of the study participants (N=7899).

Characteristics	Control notifications only (n=1985)	Affirmation quiz only (n=1984)	Notification affirmations only (n=1903)	Both quiz and notification affirmations (n=2027)	Overall (N=7899)	Test statistics		*P* value[a]
						F test (*df*)	Chi-square (*df*)	
Age (years), mean (SD)	30.5 (8.7)	30.5 (8.6)	30.5 (8.8)	30.6 (8.6)	30.5 (8.7)	0.05 (7898)	N/A[b]	.99
Gender, n (%)						N/A	1.6 (3)	.65
Male	1199 (60.4)	1225 (61.7)	1138 (59.8)	1228 (60.6)	4790 (60.6)			
Female	786 (39.6)	759 (38.3)	765 (40.2)	799 (39.4)	3109 (39.4)			
Currently using cessation aid, n (%)						N/A	4.3 (3)	.23
Yes	440 (22.2)	401 (20.2)	434 (22.8)	446 (22)	1721 (21.8)			
No	1545 (77.8)	1583 (74)	1469 (77.2)	1581 (78)	6178 (78.2)			
Time to first cigarette, nicotine dependence, n (%)						N/A	5.4 (9)	.80
Within 5 min	505 (25.4)	494 (24.9)	467 (24.5)	495 (24.4)	1961 (24.8)			
6 to 30 min	386 (19.4)	369 (18.6)	361 (19)	411 (20.3)	1527 (19.3)			
31 to 60 min	572 (28.8)	588 (29.6)	576 (30.3)	563 (27.8)	2299 (29.1)			
After 60 min	520 (26.2)	533 (26.9)	499 (26.2)	558 (27.5)	2110 (26.7)			
Missing[c]	2 (0.1)	0 (0)	0 (0)	0 (0)	2 (0.1)			
Desire to smoke[d], mean (SD)	3.1 (1.4)	3.1 (1.4)	3.2 (1.3)	3.2 (1.3)	3.2 (1.3)	3.28 (7898)	N/A	.02
Cessation stage of change, n (%)						N/A	11.7 (3)	.009
Yes, within next 30 days	1770 (89.2)	1897 (91.1)	1735 (91.2)	1795 (88.6)	7107 (90.0)			
Yes, within the next 6 months or no	215 (10.8)	177 (8.9)	168 (8.8)	232 (11.4)	792 (10.0)			
Felt happy[e], mean (SD)	3.2 (0.9)	3.3 (0.9)	3.2 (0.9)	3.2 (0.9)	3.2 (0.9)	0.73 (7898)	N/A	.54
Felt angry[e], mean (SD)	2.9 (0.9)	2.8 (0.9)	2.9 (0.9)	2.8 (0.9)	2.8 (0.9)	0.50 (7898)	N/A	.68
Felt anxious[e], mean (SD)	3.2 (1.1)	3.1 (1.1)	3.1 (1.1)	3.1 (1.1)	3.1 (1.1)	0.97 (7898)	N/A	.41
Felt hopeful[e], mean (SD)	3.1 (1.0)	3.1 (1.0)	3.1 (1.0)	3.1 (1.0)	3.1 (1.0)	0.41 (7898)	N/A	.75
Felt sad[e], mean (SD)	2.9 (1.0)	2.8 (1.0)	2.8 (1.0)	2.9 (1.0)	2.8 (1.0)	2.06 (7898)	N/A	.10
Tendency to spontaneously self-affirm[f], mean (SD)	3.1 (1.1)	3.2 (1.1)	3.1 (1.2)	3.1 (1.1)	3.1 (1.1)	2.01 (7898)	N/A	.11
Optimism[f], mean (SD)	3.6 (1.2)	3.7 (1.2)	3.6 (1.2)	3.6 (1.2)	3.6 (1.2)	2.28 (7898)	N/A	.08
Age started smoking[g], mean (SD)	17.7 (4.2)	17.6 (4.1)	17.8 (3.9)	17.7 (4.3)	17.7 (4.1)	0.32 (7836)	N/A	.81
Cigarettes per day[h], mean (SD)	14.7 (8.2)	14.6 (8.1)	14.7 (8.2)	14.7 (8.6)	14.7 (8.2)	0.14 (7882)	N/A	.94
Quit attempts in past year[i], mean (SD)	7.9 (27.5)	6.5 (20.7)	6.6 (19.6)	7.2 (22.8)	7.9 (27.5)	1.55 (7884)	N/A	.20

[a]Significance assessed using the Bonferroni-corrected α level=0.05/16 comparisons=0.003125. No variables assessed met the threshold for statistical significance after the Bonferroni correction was applied.

[b]N/A: not applicable.

[c]Respondents who were missing a valid answer for time to first cigarette were included in all subsequent analyses if they had valid data for all other variables.

[d]1=not at all; 5=a lot.

[e]1=none of the time; 5=all of the time.

[f]1=strongly disagree; 5=strongly agree.

[g]Age started smoking was a write-in question; all ages from 5 years old to present age were considered valid cases. All ages outside of this range were excluded from analysis of age started smoking but were included in all subsequent analyses.

[h]Cigarettes per day was a write-in question; all values from 0 to 99 were considered valid cases. All responses outside of this range were excluded from the analysis of cigarettes per day but were included in all subsequent analyses.

[i]Past-year cessation attempts were assessed via a write-in question; all values from 0 to 365 were considered valid cases. All responses outside of this range were excluded from analysis of past year cessation attempts but were included in all subsequent analyses.

Baseline Measures

Upon agreeing to participate, participants provided their age, gender, age at which they started smoking, average number of cigarettes smoked per day, smoking cessation aids (if any) they were currently using, and a quit date. Data on race and ethnicity were not collected, in part because the app is available globally and participants came from different countries with different racial and ethnic groups.

Participants were also asked questions to assess potential differences in baseline smoking behavior and levels of addiction. Baseline measures used to compare groups included nicotine dependence (*How soon after you wake up do you smoke your first cigarette?* [63]) and desire to smoke (*How strong is your desire to smoke, right now?* with options *not at all* to *a lot* on a 5-point scale). Previous quit attempts were assessed (*In the last year, how many times have you quit smoking for at least 24 hours?*). Smoking cessation stage of change was assessed using the following item: *Are you seriously thinking of quitting smoking?* with answer choices *yes, within the next 30 days*, corresponding to the Transtheoretical Model's preparation stage, *yes, within the next 6 months*, representing the Transtheoretical Model's contemplation stage, and *no, not thinking of quitting*, corresponding to precontemplation [64]. On the basis of the distribution of responses and our conceptual interest in the effects of self-affirmation among smokers who intend to quit smoking, the stage of change was dichotomized into *yes, within the next 30 days* and *yes, within the next 6 months* or *no*. These items were used to compare groups at baseline for smoking behavior and experiences.

Affect was assessed using items from the Positive and Negative Affect Schedule (PANAS) [65] as adapted by the Midlife in the United States (MIDUS) study [66,67] and the National Cancer Institute (NCI) Health Information National Trends Survey (HINTS). Participants rated their level of happiness, anger, anxiety, sadness, hopefulness, and anxiety within the past 30 days on a 5-point scale from *none of the time* to *all of the time*. The affect items were reverse coded so that higher scores indicated experiencing that emotion more often. Participants' tendencies to engage in spontaneous self-affirmation were assessed as the average of 2 items used in previous studies [45,48] from the longer spontaneous self-affirmation measure (SSAM [41]): (1) "When I feel threatened or anxious I find myself thinking about my strengths" and (2) "When I feel threatened or anxious I find myself thinking about my values." Participants' baseline level of optimism was assessed with the item: "I'm always optimistic about my future" [68]. The SSAM and optimism items were assessed on a 5-point scale with the anchors 1=strongly agree to 5=strongly disagree. However, these items were reverse coded so that higher scores indicated more agreement and thus higher optimism.

Follow-Up Surveys (1 and 3 Months)

All participants were invited to complete a 1-month and 3-month follow-up survey to assess smoking status. Both follow-up surveys assessed smoking status with the items: "Have you smoked at all in the past month?" and "Have you smoked at all in the past week?" Response options for both questions were "no, not a puff," "1-5 cigarettes," or "more than 5 cigarettes." For these analyses, responses were dichotomized to indicate no smoking or smoking (1 cigarette or more) in the time period. In addition, participants were asked to report the average number of cigarettes smoked per day, time to first cigarette, and if they were using any other cessation aids at follow-up.

Baseline Self-Affirmation Questionnaire

Participants assigned to the baseline affirmation conditions were shown a shortened, 5-item kindness questionnaire (quiz), adapted from previous work [61,69], directly after the baseline questionnaire. The purpose of this quiz was to induce self-affirmation by allowing participants to respond *yes* to having engaged in specific instances of past kindness. In the original questionnaire, participants were asked to provide written examples for each item to which they responded affirmatively; however, participants in this study were not asked to provide examples. This self-affirmation induction has been frequently used, has face validity, and is easy to implement [70]. The control condition did not receive the kindness quiz or any content in its place. The full self-affirmation questionnaire as well as responses by condition is presented in Multimedia Appendix 2. Of the respondents who received the baseline kindness quiz, approximately 83.1% (3333/4011) answered *yes* to 4 of the 5 items (Multimedia Appendix 2).

Affirmation and Control Push Notifications

Participants in the control push notification condition received general tips related to quitting smoking, whereas participants in the affirmation push notifications condition received affirming messages from a pool of 15 possible notifications. The affirmation messages were based on literature and a previous study of self-affirmation content that had been integrated into a text messaging intervention for smoking cessation [61]. The control notifications were informed by the smoking cessation

literature [71,72]. Participants received 2 notifications (either self-affirmation or control, depending on their assigned condition) per day for the duration of the study, unless they turned off notifications, which was the same as the frequency of notifications in the current version of the app. One notification was sent during each of the following time blocks: 8 AM to 2:30 PM and 2:31 PM to 9 PM. Participants were able to change the earliest and latest time for the notification (eg, change 8 AM to 7 AM or 9 PM to 11 PM). Participants could also access these self-affirmation or control messages (depending on condition) every time they reported experiencing a craving on a *Tips* screen. The full text of all notifications, organized by category, is presented in Multimedia Appendix 3.

Analysis

All analyses were conducted using Stata 16 [73]. First, attrition rates were calculated for each of the 4 induced self-affirmation conditions. Second, demographic characteristics and baseline survey responses were compared across groups using ANOVA and chi-square tests. Third, a series of binary logistic regression models were run to examine predictors of successful cessation and potential moderating factors. All regression analyses used intent-to-treat, such that respondents who did not provide follow-up data were treated as continuing smokers. We elected to use intent-to-treat because it is widely used for assessing smoking cessation in interventions [8,9] and tends to be more conservative in assuming that all participants lost to follow-up continued to smoke instead of artificially inflating the cessation rate by removing participants lost to follow-up from analyses. We adopted P=.05 as our cut-off for statistical significance, with Bonferroni corrections applied for multiple comparisons as necessary.

Trial Registration

The trial was retrospectively registered at ISRCTN: https://www.isrctn.com/ISRCTN56646695.

Results

Enrollment, Attrition, and Participant Characteristics

A CONSORT (Consolidated Standards of Reporting Trials) diagram is provided to show subject enrollment and study completion (Figure 1). Overall, 8.2% (647/7899) of users who enrolled in the study completed the 1-month follow-up survey and 2.9% (231/7899) completed the 3-month survey (Table 2), consistent with systematic reviews that found high levels of attrition in web-based RCTs [74]. The proportion of users who completed the follow-up was lower than in previous studies of this same app in which 7.5% of participants completed a 3-month follow-up [15]; however, the sample size in this study was considerably smaller. Attrition in this study was similar to attrition in a similar previous study in which 6.4% of participants completed a 42-day follow-up [61]. Although we estimated 90% attrition in our power calculations, we exceeded this percentage. In addition, there were more follow-up assessments in this study and it was available to Android but not iOS users, in contrast with the previous study [15], which had fewer and shorter follow-up assessments and enrolled both iOS and Android users. It is possible that the more frequent follow-ups combined with differences between the iOS and Android app can help contextualize this finding. Follow-up rates did not differ significantly by condition (1 month: χ^2_3=0.6, P=.90; 3 months: χ^2_3=1.5, P=.68; Table 2). This paper presents analyses for both the 1- and 3-month follow-ups; however, as the 3-month follow-up rates were considerably lower than the 1-month follow-up rates, most implications and conclusions focus on results from the 1-month follow-up.

Table 2. Attrition and cessation rates by study condition (N=7899).

Outcome	Control notifications only	Affirmation quiz only	Notification affirmations only	Both quiz and notification affirmations	Overall	Test statistic, chi square (*df*)	*P* value
Completed baseline survey, n	1985	1984	1903	2027	7899	—[a]	—
Completed 1-month follow-up, n (%)	164 (8.3)	165 (8.3)	160 (8.4)	158 (7.8)	647 (8.2)	0.6 (3)	.90
Past week cessation at 1 month	145 (7.3)	143 (7.2)	139 (7.3)	142 (7.0)	569 (7.2)	0.2 (3)	.98
Past month cessation at 1 month	133 (6.7)	128 (6.5)	128 (6.7)	114 (5.6)	503 (6.4)	2.7 (3)	.44
Completed 3-month follow-up, n (%)	51 (2.6)	64 (3.2)	56 (2.9)	60 (3.0)	231 (2.9)	1.5 (3)	.68
Past week cessation at 3 months	48 (2.4)	58 (2.9)	55 (2.9)	54 (2.7)	215 (2.7)	1.2 (3)	.75
Past month cessation at 3 months	38 (1.9)	49 (2.5)	48 (2.5)	46 (2.3)	181 (2.3)	2.0 (3)	.57

[a]No statistical tests were run.

Smoking Cessation and Baseline Differences

Cessation rates did not differ significantly between groups at the 1-month (past-week cessation: χ^2_3=0.2, P=.98; past-month cessation: χ^2_3=2.7, P=.44; Table 2) or 3-month (past week cessation: χ^2_3=1.2, P=.75; past month cessation: χ^2_3=2.0, P=.57; Table 2) follow-up. Using an intent-to-treat analysis, the overall past-week cessation rate was 7.2% (569/7899) and the past-month cessation rate was 6.4% (503/7899) at the 1-month follow-up (Table 2). This is similar to the previous study of a text messaging program with affirmation content, which found 5.6% cessation at 6 weeks using intent-to-treat analysis [61]. Notably, despite randomization, participants differed in baseline

desire to smoke (F_{7898}=3.28; *P*=.02; Table 1) and cessation stage of change (χ^2_3=11.7; *P*=.009; Table 1) across conditions, although neither met the threshold for statistical significance after the Bonferroni correction was applied. The Bonferroni correction is conservative; because cessation stage of change differed at the *P*<.01 level and is likely related to the smoking cessation outcome, subsequent regression analyses controlled for baseline cessation stage of change.

Aims 1 and 2: Self-Affirmation's Associations With Cessation Outcomes

The primary aim of this study was to assess the impact of induced self-affirmation conditions on smoking cessation. The secondary aim of this study was to assess the associations of spontaneous self-affirmation with smoking cessation. To assess factors associated with cessation, binary logistic regression models were run using one of the 2 main outcomes (past-week cessation at 1 month and past-month cessation at 1 month) as the dependent variable. The 2 main 1-month outcomes were strongly correlated (*r*=0.93). In both regression models, tendency to spontaneously self-affirm at baseline was a significant predictor of cessation (Table 3), consistent with the hypotheses. However, neither self-affirmation study condition nor their interaction was significant in these models, indicating that providing opportunities for self-affirmation in the smoking cessation smartphone app did not result in a greater likelihood of cessation than using the smartphone app without affirmation.

Table 3. Primary self-affirmation regression models (N=7899).

Variable	Past-week cessation at 1 month			Past-month cessation at 1 month		
	OR[a] (95% CI)	SE	*P* value	OR (95% CI)	SE	*P* value
Baseline affirmation	0.99 (0.78-1.26)	0.1	.92	0.96 (0.75-1.24)	0.1	.76
Notification affirmations	0.99 (0.77-1.26)	0.1	.92	0.99 (0.77-1.28)	0.1	.97
Baseline and notification affirmations interaction	0.98 (0.70-1.38)	0.2	.90	0.87 (0.60-1.12)	0.2	.44
Cessation stage of change: Yes, within the next 6 months or No[b]	0.80 (0.59-1.10)	0.1	.17	0.85 (0.62-1.18)	0.1	.40
Spontaneous self-affirmation	0.85 (0.79-0.92)	0.0	<.001[c]	0.90 (0.83-0.97)	0.0	.01[d]

[a]OR: odds ratio.

[b]Reference category: Yes, within the next 30 days.

[c]*P*<.001.

[d]*P*<.05.

In addition, despite low follow-up rates, both 3-month outcomes were also explored. There was a similarly high association between past-week and past-month cessation at the 3-month follow-up (*r*=0.91). None of the self-affirmation measures (tendency to spontaneously self-affirm, baseline affirmation condition, notification affirmation condition, the interaction of baseline, and notification affirmations) significantly predicted cessation in the main 3-month models (Multimedia Appendix 4).

Subsequently, we ran models testing a three-way interaction between our self-affirmation study conditions and spontaneous self-affirmation with both the 1-month and 3-month outcomes. This interaction was not significant, establishing that spontaneous self-affirmation did predict 1-month cessation in this study, but did not moderate the relationship between study conditions and cessation outcomes.

Aim 3: Examining Potential Predictors and Moderators

The goal of exploratory aim 3 was to assess baseline optimism and baseline affective states as potential predictors and moderators of the relationship between affirmation conditions and cessation outcomes. To test this aim, we conducted regression analyses in which optimism and each of the 5 affective states (happiness, anger, anxiousness, hopefulness, sadness) were simultaneously added to the main regression models (Table 4). That is, we tested whether these factors predicted 1-week and 1-month cessation at 1 month when controlling for spontaneous self-affirmation, baseline affirmation condition, notification affirmation condition, their interaction, and cessation stage of change. In this model, tendency to spontaneously self-affirm still significantly predicted past-week cessation at the 1-month follow-up (*P*<.001) but did not significantly predict past-month cessation at the 1-month follow-up, although this association approached significance (*P*=.05). Lower sadness was a significant predictor of successful cessation at the 1-month follow-up for both past-week and past-month cessation (*P*=.002 and *P*=.007, respectively). Optimism predicted past-week cessation at the 1-month follow-up (*P*=.04) and both happiness and anger predicted past-month cessation at the 1-month follow-up (*P*=.03 for both). We also ran these with each potential predictor assessed separately and the results did not differ from the simultaneous model presented in Table 4.

Table 4. Regression models with potential predictors (N=7899).

Variable	Past-week cessation at 1 month			Past-month cessation at 1 month		
	OR[a] (95% CI)	SE	*P* value	OR (95% CI)	SE	*P* value
Baseline affirmation	0.97 (0.77-1.24)	0.1	.83	0.95 (0.74-1.22)	0.1	.68
Notification affirmations	0.98 (0.77-1.25)	0.1	.87	0.99 (0.77-1.27)	0.1	.92
Baseline and notification affirmations interaction	1.00 (0.71-1.40)	0.2	.98	0.88 (0.61-1.26)	0.2	.48
Cessation stage of change: Yes, within the next 6 months or No[b]	0.81 (0.59-1.11)	0.1	.19	0.86 (0.63-1.19)	0.1	.37
Spontaneous self-affirmation	0.86 (0.79-0.94)	0.0	<.001[c]	0.91 (0.84-1.00)	0.0	.05
Optimism	0.92 (0.85-0.99)	0.0	.04[d]	0.93 (0.85-1.00)	0.0	.09
Happy	1.05 (0.93-1.18)	0.1	.42	0.91 (0.83-0.99)	0.0	.03[d]
Angry	0.90 (0.80-1.01)	0.1	.06	0.87 (0.78-0.99)	0.1	.03[d]
Anxious	0.97 (0.88-1.06)	0.1	.48	0.97 (0.88-1.08)	0.1	.61
Hopeful	1.04 (0.94-1.14)	0.1	.47	1.03 (0.93-1.14)	0.1	.56
Sad	0.84 (0.75-0.93)	0.1	.002[d]	0.85 (0.76-0.96)	0.1	.007[d]

[a]OR: odds ratio.

[b]Reference category: Yes, within the next 30 days.

[c]*P*<.001.

[d]*P*<.05.

Next, moderation analyses were conducted, with a three-way interaction term (baseline affirmation, notification affirmation, and potential moderator) for each of the affective states and optimism, run separately. No three-way interaction was significant. The full model for sadness is provided as an appendix (Multimedia Appendix 5); however, the models for other affective states and optimism have not been included because of space limitations.

We then conducted parallel analyses with the 3-month outcomes. None of the affective states were significant predictors of cessation in the past week or month at the 3-month follow-up; however, optimism was a predictor of cessation for both past-week and past-month cessation at the 3-month follow-up (*P*=.004 and *P*=.002, respectively; Multimedia Appendix 4). Due to low follow-up rates at the 3-month follow-up, potential moderation was not explored with the 3-month outcomes.

Association of Spontaneous Self-Affirmation With Dependence

Previous work has found that smokers with a greater tendency to spontaneously self-affirm report more quit attempts and higher quit intentions, particularly when they live in states with more comprehensive smoke-free laws, highlighting factors that affect the cessation process [48]. Due to the significant effects of spontaneous self-affirmation in our study, we undertook additional analyses. The mean spontaneous self-affirmation scores in our study (mean 3.11 out of 5, SD 1.1) were comparable with those of past studies (mean 2.75 out of 4, SD 0.14 [46]; mean 3.12 out of 5, SD 0.86 [75]).

We examined whether spontaneous self-affirmation was associated with several dependence and quit intention measures in our sample to assess whether respondents who had a greater tendency to spontaneously self-affirm were less dependent on nicotine or had a stronger desire to quit at baseline, offering them an advantage. We computed correlations of spontaneous self-affirmation with the cessation stage of change, time to first cigarette, and quit intention. All were small (all correlation coefficient values were less than 0.1), indicating that smokers higher in tendency to self-affirm in this study were not necessarily less addicted or more intent to quit at baseline than smokers with lower tendencies to self-affirm.

Discussion

Principal Findings

In this study, opportunities for self-affirmation provided in the smartphone app (baseline self-affirmation quiz and self-affirmation notifications) did not significantly improve the likelihood of successful cessation. However, tendency to spontaneously self-affirm was a strong and significant predictor of cessation. Baseline sadness was associated with a lower likelihood of reporting successful cessation at the 1-month follow-up; optimism was significantly associated with past-week cessation at the 1-month follow-up, and happiness and anger were both significantly associated with past-month cessation at the 1-month follow-up. There were no interactions between any explored individual difference predictor and study conditions.

The spontaneous self-affirmation findings are consistent with previous findings that spontaneous self-affirmation was associated with improved psychological well-being and health care experiences [45,46], both of which may play a role in the smoking cessation process. In addition, although previous studies have found a relationship between spontaneous self-affirmation and quit attempts and intentions [48], we did not find a relationship between spontaneous self-affirmation

and any dependence or quit measure in our sample. Thus, any association of spontaneous self-affirmation with quitting was not due to dependence or past quit attempts. Future work should explore this relationship in more detail to understand the specific benefit and ways to help smokers who do not tend to spontaneously self-affirm.

Comparison With Previous Work

This study differs from previous smoking self-affirmation studies that typically provide participants with information about the negative health consequences of smoking. In this study, no explicitly threatening health information or loss-framed messages were provided, consistent with the positive focus of the app. The messages conveyed the benefits of quitting instead of the harms of smoking and were not hypothesized to constitute explicitly threatening health information. We did not directly assess whether participants perceived any of the material in the app to be threatening. One recent meta-analysis found that self-affirmation is less likely to facilitate change when psychological threat is minimal [32], which suggests that self-affirmation opportunities may have been less effective in this study in the absence of directly threatening information. Moreover, smokers may be aware of the health costs of smoking, and it is unknown to what extent information about the health consequences is novel to smokers, which could help to explain null self-affirmation effects in previous studies of smokers. These studies would benefit from pilot testing to determine whether threatening information about the health consequences of smoking is indeed perceived by smokers as threatening and novel. An alternative explanation is that our control messages were received well by respondents, offsetting our ability to observe any benefit of the self-affirmation messages.

Previous studies have reported mixed findings concerning whether self-affirmation inductions can assist smokers trying to quit; some studies have found benefits [28,30,33-35], whereas others have not [35,39,40,76]. Our finding that induced self-affirmations did not influence smoking behavior is consistent with multiple other studies that have shown null or even backfiring effects among smokers who undergo self-affirmation interventions [35]. It is possible that our baseline affirmation quiz and notification affirmations did not induce self-affirmation in participants; consistent with previous self-affirmation intervention studies, no manipulation check for affirmation was included, and it is difficult to assess whether participants were successfully affirmed. The original kindness quiz asks respondents to write down a specific time they engaged in the aforementioned action [69]. In this study, the baseline kindness quiz was adapted to ask participants to answer *yes* or *no* without explicitly asking them to write or think of an example, given that the affirmation intervention occurred through text messages.

It is a challenge to determine how best to adapt self-affirmation interventions developed in laboratory settings to the real world. In a previous study testing whether various adaptations of the kindness quiz differentially affected health cognitions and smoking intentions among a sample of online smokers, there were no significant differences depending on whether participants were asked to write examples, imagine examples, or were not asked to provide any examples [35]. However, that study also found that none of the self-affirmation conditions were more effective than the control conditions [35]. In addition, participants asked to provide written self-affirmation responses endorsed fewer affirmation questions than those not asked to provide written examples, suggesting that the writing was onerous [35]. In that study, for participants providing written examples, the intervention took nearly 7 times longer than it did for participants not asked to provide examples [35]. Thus, more research is needed to determine how to administer effective self-affirmation interventions to participants not in laboratory settings.

Our finding that additional opportunities for self-affirmation added to the smartphone app in this study did not have effects could be due to multiple factors. The self-affirmation content may not have been as noticeable as self-affirmations in other studies, given that participants did not complete the study in a more controlled laboratory setting. In addition, participants may have skimmed or otherwise not engaged with the self-affirmation content in this study. We also do not have data on the extent to which participants read or engage with induced self-affirmation materials. In addition, the existing app material was evidence-based and has already been found to be relatively effective on its own [15], thus identifying additional benefits of novel self-affirmation intervention material may have been difficult.

This study complements existing evidence concerning the distinctiveness of spontaneous self-affirmation from other psychological resources, such as optimism [41]. In this study, the single-item measure of optimism was only moderately correlated with spontaneous self-affirmation (r=0.46). Some previous work using a small number of items has found that spontaneous self-affirmation is related to greater optimism [46]. However, the correlation between the full measure of spontaneous self-affirmation and optimism is small (eg, r=0.22 as observed in a study by Harris et al [41]). Furthermore, cancer survivors who reported greater optimism reported better physical, mental, and cognitive health, even when controlling for spontaneous self-affirmation [77]. In this study, spontaneous self-affirmation was a significant predictor of 1-month cessation outcomes, whereas optimism was unrelated to 1-month cessation outcomes but predicted 3-month cessation outcomes. Optimism facilitates pursuit of goals [50,51], so it is noteworthy that it was not associated with smoking cessation goals at 1 month, whereas spontaneous self-affirmation did maintain such an association.

Interestingly, baseline sadness was significantly related to cessation outcomes at the 1-month follow-up. Feeling sadness less frequently at baseline was associated with a greater likelihood of reporting both past-week and past-month cessation. The relationship between sadness and cessation outcomes is consistent with previous theory and research suggesting that sadness facilitates reward-seeking tendencies that might undermine healthy behavior, including smoking cessation [55-57]. Optimism and all affective states (happiness, anger, anxiousness, hopefulness, sadness) were not found to moderate the relationship between the assigned affirmation conditions and successful cessation. Previous work has found that clinical

diagnoses of anhedonia and depressed mood predict increased odds of relapse among smokers trying to quit [78]; however, this study is among the first to examine specific affective states and their association with successful cessation. Future work can further disentangle the relationship between sadness and cessation experiences.

Limitations

This study has several limitations. As previously discussed, we do not have data concerning whether participants were successfully affirmed and to what extent they were engaged by the intervention. There are several methodological limitations. As data were collected from all users who downloaded the smartphone app, it was difficult to maintain strict experimental control. We were not able to monitor if or when participants turned off notifications, so we were unable to assess an individual's exposure to the notification content. We were also unable to determine the geographic location of individual participants and were only able to access aggregate geographic information for the sample. Participants came largely from the United Kingdom and the United States, but there were participants from 8 other countries. In addition, smoking cessation is a complex process, and whereas many users who completed the baseline assessments did not complete follow-up assessments, we were not able to analytically determine why these participants discontinued responding and if they had deleted the smartphone app due to successful cessation or another reason. Another limitation is the use of 1 or 2-item measures of key constructs, such as spontaneous self-affirmation. However, these items have shown significant associations with outcomes in other studies [41], thus providing support for their validity. Similarly, we only assessed affect once at baseline. Emotions fluctuate over time, particularly during the difficult smoking cessation process. Future studies can monitor changes in affect during the process, such as with daily dairies, to better understand the role that affects plays in cessation. Finally, the attrition experienced in this study was higher than expected based on previous similar studies [15,61,74]. In this study, we found that 8.2% (647/7899) of users who enrolled in the study completed the 1-month follow-up survey and 2.9% (231/7899) completed the 3-month survey, which is lower than the 7.5% of participants who completed a 3-month follow-up during a previous trial of this same app [15]. In a previous study that informed the present study, 6.4% of participants completed a 42-day follow-up [61]. High attrition limits the interpretability of results such that it may have made it difficult to detect and reduce the generalizability of results, particularly at the 3-month follow-up. Furthermore, the intent-to-treat approach assumes that nonresponders are smokers, whereas it could be the case that nonresponders found the protocol burdensome.

However, these limitations are offset by several considerable strengths of this study. This study used a sample of real-life users, which allows for an assessment of how the app will function outside of a highly controlled laboratory setting. The study was also theoretically driven and provides preliminary evidence for the promise of spontaneous self-affirmation in smoking cessation. An additional strength of this study comes from the use of an already-existing, successful smoking cessation app with the addition of self-affirmation specific content.

Conclusions

The results of this study provide evidence that spontaneous self-affirmation may be an important threat management psychological resource in the context of smoking cessation. They indicate the difficulties of creating effective self-affirmation inductions in smoking apps. There is a need to examine the effectiveness of smartphone app–delivered self-affirmations and to develop more effective affirmations in future dissemination work.

Conflicts of Interest

DC is originator of the Smoke Free app and derives income from it.

Multimedia Appendix 1
Geographic locations of the participants.
[DOCX File , 62 KB - jmir_v23i3e18433_app1.docx]

Multimedia Appendix 2
Baseline self-affirmation questionnaire and responses by condition.
[DOCX File , 34 KB - jmir_v23i3e18433_app2.docx]

Multimedia Appendix 3
Text notifications by study condition.
[DOCX File , 35 KB - jmir_v23i3e18433_app3.docx]

Multimedia Appendix 4
Regression models for 3-month outcomes.
[DOCX File , 41 KB - jmir_v23i3e18433_app4.docx]

Multimedia Appendix 5
Regression models to explore sadness as a potential moderator of self-affirmation conditions.
[DOCX File , 37 KB - jmir_v23i3e18433_app5.docx]

Multimedia Appendix 6
CONSORT-EHEALTH (V.1.6) checklist.
[PDF File (Adobe PDF File), 1620 KB - jmir_v23i3e18433_app6.pdf]

References

1. Drope J, Schluger N, Cahn Z, Drope J, Hamill S, Islami F, et al. Chapter 8: Deaths. In: The Tobacco Atlas. Atlanta, GA: American Cancer Society; 2018:28.
2. Henley SJ, Thomas CC, Sharapova SR, Momin B, Massetti GM, Winn DM, et al. Vital signs: disparities in tobacco-related cancer incidence and mortality — United States, 2004–2013. MMWR Morb Mortal Wkly Rep 2016 Nov 11;65(44):1212-1218. [doi: 10.15585/mmwr.mm6544a3]
3. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting smoking among ddults - United States, 2000-2015. MMWR Morb Mortal Wkly Rep 2017 Jan 06;65(52):1457-1464 [FREE Full text] [doi: 10.15585/mmwr.mm6552a1] [Medline: 28056007]
4. Centers for Disease ControlPrevention (CDC). Quitting smoking among adults--United States, 2001-2010. MMWR Morb Mortal Wkly Rep 2011 Nov 11;60(44):1513-1519 [FREE Full text] [Medline: 22071589]
5. Fiore M, Jaén C, Baker T. Evidence and recommendations, treatment elements. In: Treating Tobacco Use and Dependence: 2008 Update. Clinical Practice Guideline. Rockville, MD: US Department of Health and Human Services. Public Health Service; May 2008:100-103.
6. Lancaster T, Stead L. Individual behavioural counselling for smoking cessation. Cochrane Database Syst Rev 2017 Mar 31(3):1-68. [doi: 10.1002/14651858.cd001292.pub3] [Medline: 28361496]
7. Piasecki TM. Relapse to smoking. Clin Psychol Rev 2006 Mar;26(2):196-215. [doi: 10.1016/j.cpr.2005.11.007] [Medline: 16352382]
8. Fraser D, Kobinsky K, Smith SS, Kramer J, Theobald WE, Baker TB. Five population-based interventions for smoking cessation: a MOST trial. Transl Behav Med 2014 Dec;4(4):382-390. [doi: 10.1007/s13142-014-0278-8] [Medline: 25584087]
9. Spohr SA, Nandy R, Gandhiraj D, Vemulapalli A, Anne S, Walters ST. Efficacy of SMS text message interventions for smoking cessation: a meta-analysis. J Subst Abuse Treat 2015 Sep;56:1-10. [doi: 10.1016/j.jsat.2015.01.011] [Medline: 25720333]
10. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:1-52 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
11. Mobile Fact Sheet. Pew Research Center. 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-12-22]
12. Fox, S. The social life of health information. Pew Research Center. 2014 Jan 15. URL: https://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/ [accessed 2020-12-22]
13. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
14. Ubhi HK, Kotz D, Michie S, van Schayck OC, Sheard D, Selladurai A, et al. Comparative analysis of smoking cessation smartphone applications available in 2012 versus 2014. Addict Behav 2016 Jul;58:175-181 [FREE Full text] [doi: 10.1016/j.addbeh.2016.02.026] [Medline: 26950256]
15. Crane D, Ubhi HK, Brown J, West R. Relative effectiveness of a full versus reduced version of the 'Smoke Free' mobile application for smoking cessation: an exploratory randomised controlled trial. F1000Res 2018;7:1524 [FREE Full text] [doi: 10.12688/f1000research.16148.2] [Medline: 30728950]
16. US Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2014:1-943.
17. Minami H, Yeh VM, Bold KW, Chapman GB, McCarthy DE. Relations among affect, abstinence motivation and confidence, and daily smoking lapse risk. Psychol Addict Behav 2014 Jun;28(2):376-388 [FREE Full text] [doi: 10.1037/a0034445] [Medline: 24955665]
18. Gwaltney CJ, Metrik J, Kahler CW, Shiffman S. Self-efficacy and smoking cessation: a meta-analysis. Psychol Addict Behav 2009 Mar;23(1):56-66 [FREE Full text] [doi: 10.1037/a0013529] [Medline: 19290690]
19. Steele C. The psychology of self-affirmation: sustaining the integrity of the self. Adv Exp Soc Psychol 1988;21:261-302. [doi: 10.1016/S0065-2601(08)60229-4]
20. Liberman A, Chaiken S. Defensive processing of personally relevant health messages. Pers Soc Psychol Bull 2016 Jul 02;18(6):669-679. [doi: 10.1177/0146167292186002]

21. van ‘t Riet J, Ruiter RA. Defensive reactions to health-promoting information: an overview and implications for future research. Health Psychol Rev 2013 May;7(sup1):S104-S136. [doi: 10.1080/17437199.2011.606782]
22. Jemmott JB, Ditto PH, Croyle RT. Judging health status: effects of perceived prevalence and personal relevance. J Pers Soc Psychol 1986 May;50(5):899-905. [doi: 10.1037//0022-3514.50.5.899] [Medline: 3712230]
23. Kunda Z. Motivated inference: self-serving generation and evaluation of causal theories. J Pers Soc Psychol 1987;53(4):636-647. [doi: 10.1037/0022-3514.53.4.636]
24. Erceg-Hurn D, Steed L. Does exposure to cigarette health warnings elicit psychological reactance in smokers? J Appl Soc Psychol 2011;41(1):219-237. [doi: 10.1111/j.1559-1816.2010.00710.x]
25. Hall MG, Sheeran P, Noar SM, Ribisl KM, Bach LE, Brewer NT. Reactance to health warnings scale: development and validation. Ann Behav Med 2016 Oct;50(5):736-750 [FREE Full text] [doi: 10.1007/s12160-016-9799-3] [Medline: 27333895]
26. Oakes W, Chapman S, Borland R, Balmford J, Trotter L. 'Bulletproof skeptics in life's jungle': which self-exempting beliefs about smoking most predict lack of progression towards quitting? Prev Med 2004 Oct;39(4):776-782. [doi: 10.1016/j.ypmed.2004.03.001] [Medline: 15351545]
27. Sherman DA, Nelson LD, Steele CM. Do messages about health risks threaten the self? Increasing the acceptance of threatening health messages via self-affirmation. Pers Soc Psychol Bull 2016 Jul 02;26(9):1046-1058. [doi: 10.1177/01461672002611003]
28. Harris PR, Mayle K, Mabbott L, Napper L. Self-affirmation reduces smokers' defensiveness to graphic on-pack cigarette warning labels. Health Psychol 2007;26(4):437-446. [doi: 10.1037/0278-6133.26.4.437]
29. Klein WM, Lipkus IM, Scholl SM, McQueen A, Cerully JL, Harris PR. Self-affirmation moderates effects of unrealistic optimism and pessimism on reactions to tailored risk feedback. Psychol Health 2010 Dec;25(10):1195-1208 [FREE Full text] [doi: 10.1080/08870440903261970] [Medline: 20204982]
30. Epton T, Harris PR, Kane R, van Koningsbruggen GM, Sheeran P. The impact of self-affirmation on health-behavior change: a meta-analysis. Health Psychol 2015 Mar;34(3):187-196. [doi: 10.1037/hea0000116] [Medline: 25133846]
31. Sweeney AM, Moyer A. Self-affirmation and responses to health messages: a meta-analysis on intentions and behavior. Health Psychol 2015 Feb;34(2):149-159. [doi: 10.1037/hea0000110] [Medline: 25089345]
32. Ferrer RA, Cohen GL. Reconceptualizing self-affirmation with the trigger and channel framework: lessons from the health domain. Pers Soc Psychol Rev 2019 Aug;23(3):285-304. [doi: 10.1177/1088868318797036] [Medline: 30295141]
33. Armitage CJ, Harris PR, Hepton G, Napper L. Self-affirmation increases acceptance of health-risk information among UK adult smokers with low socioeconomic status. Psychol Addict Behav 2008 Mar;22(1):88-95. [doi: 10.1037/0893-164X.22.1.88] [Medline: 18298234]
34. Pietersma S, Dijkstra A. Cognitive self-affirmation inclination: an individual difference in dealing with self-threats. Br J Soc Psychol 2012 Mar;51(1):33-51. [doi: 10.1348/014466610X533768] [Medline: 22435845]
35. Taber JM, McQueen A, Simonovic N, Waters EA. Adapting a self-affirmation intervention for use in a mobile application for smokers. J Behav Med 2019 Dec;42(6):1050-1061. [doi: 10.1007/s10865-019-00028-1] [Medline: 30903442]
36. Memish KE, Schüz N, Frandsen M, Ferguson SG, Schüz B. Using self-affirmation to increase the effects of emotive health warnings on smoking: a randomized exploratory trial. Nicotine Tob Res 2017 Oct 01;19(10):1238-1242. [doi: 10.1093/ntr/ntw167] [Medline: 27613906]
37. Bommelé J, Schoenmakers TM, Kleinjan M, Peters GY, Dijkstra A, van de Mheen D. Targeting hardcore smokers: the effects of an online tailored intervention, based on motivational interviewing techniques. Br J Health Psychol 2017 Dec;22(3):644-660. [doi: 10.1111/bjhp.12256] [Medline: 28703470]
38. Epton T, Norman P, Dadzie A, Harris PR, Webb TL, Sheeran P, et al. A theory-based online health behaviour intervention for new university students (U@Uni): results from a randomised controlled trial. BMC Public Health 2014 Jun 05;14:563 [FREE Full text] [doi: 10.1186/1471-2458-14-563] [Medline: 24903620]
39. Dillard A, McCaul K, Magnan R. Why is such a smart person like you smoking? Using self-affirmation to reduce defensiveness to cigarette warning labels. J Appl Biobehav Res 2005;10(3):165-182. [doi: 10.1111/j.1751-9861.2005.tb00010.x]
40. Zhao X, Peterson EB, Kim W, Rolfe-Redding J. Effects of self-affirmation on daily versus occasional smokers’ responses to graphic warning labels. Commun Res 2012 Nov 09;41(8):1137-1158. [doi: 10.1177/0093650212465433]
41. Harris PR, Griffin DW, Napper LE, Bond R, Schüz B, Stride C, et al. Individual differences in self-affirmation: distinguishing self-affirmation from positive self-regard. Self Identity 2018 Aug 21;18(6):589-630. [doi: 10.1080/15298868.2018.1504819]
42. Cornil Y, Chandon P. From fan to fat? Vicarious losing increases unhealthy eating, but self-affirmation is an effective remedy. Psychol Sci 2013 Oct;24(10):1936-1946. [doi: 10.1177/0956797613481232] [Medline: 23925307]
43. Toma CL, Hancock JT. Self-affirmation underlies Facebook use. Pers Soc Psychol Bull 2013 Mar;39(3):321-331. [doi: 10.1177/0146167212474694] [Medline: 23359086]
44. Ferrer RA, Taber JM, Klein WM, Harris PR, Lewis KL, Biesecker LG. The role of current affect, anticipated affect and spontaneous self-affirmation in decisions to receive self-threatening genetic risk information. Cogn Emot 2015;29(8):1456-1465 [FREE Full text] [doi: 10.1080/02699931.2014.985188] [Medline: 25482843]

45. Taber JM, Howell JL, Emanuel AS, Klein WM, Ferrer RA, Harris PR. Associations of spontaneous self-affirmation with health care experiences and health information seeking in a national survey of US adults. Psychol Health 2016;31(3):292-309 [FREE Full text] [doi: 10.1080/08870446.2015.1085986] [Medline: 26315683]
46. Emanuel AS, Howell JL, Taber JM, Ferrer RA, Klein WM, Harris PR. Spontaneous self-affirmation is associated with psychological well-being: evidence from a US national adult survey sample. J Health Psychol 2018 Jan;23(1):95-102. [doi: 10.1177/1359105316643595] [Medline: 27160152]
47. Creswell JD, Lam S, Stanton AL, Taylor SE, Bower JE, Sherman DK. Does self-affirmation, cognitive processing, or discovery of meaning explain cancer-related health benefits of expressive writing? Pers Soc Psychol Bull 2007 Feb;33(2):238-250. [doi: 10.1177/0146167206294412] [Medline: 17259584]
48. Persoskie A, Ferrer RA, Taber JM, Klein WMP, Parascandola M, Harris PR. Smoke-free air laws and quit attempts: evidence for a moderating role of spontaneous self-affirmation. Soc Sci Med 2015 Sep;141:46-55. [doi: 10.1016/j.socscimed.2015.07.015] [Medline: 26246033]
49. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. J Pers Soc Psychol 1994 Dec;67(6):1063-1078. [doi: 10.1037//0022-3514.67.6.1063] [Medline: 7815302]
50. Monzani D, Steca P, Greco A, D'Addario M, Pancani L, Cappelletti E. Effective pursuit of personal goals: the fostering effect of dispositional optimism on goal commitment and goal progress. Pers Individ Dif 2015 Aug;82:203-214. [doi: 10.1016/j.paid.2015.03.019]
51. Segerstrom SC, Nes LS. When goals conflict but people prosper: the case of dispositional optimism. J Res Pers 2006 Oct;40(5):675-693 [FREE Full text] [doi: 10.1016/j.jrp.2005.08.001] [Medline: 17016545]
52. Lerner JS, Han S, Keltner D. Feelings and consumer decision making: extending the appraisal-tendency framework. J Consum Psychol 2007 Jul;17(3):181-187. [doi: 10.1016/s1057-7408(07)70027-x]
53. Lerner JS, Keltner D. Beyond valence: toward a model of emotion-specific influences on judgement and choice. Cogn Emot 2000 Jul;14(4):473-493. [doi: 10.1080/026999300402763]
54. Lerner JS, Keltner D. Fear, anger, and risk. J Pers Soc Psychol 2001 Jul;81(1):146-159. [doi: 10.1037//0022-3514.81.1.146] [Medline: 11474720]
55. Dorison CA, Wang K, Rees VW, Kawachi I, Ericson KMM, Lerner JS. Sadness, but not all negative emotions, heightens addictive substance use. Proc Natl Acad Sci U S A 2020 Jan 14;117(2):943-949 [FREE Full text] [doi: 10.1073/pnas.1909888116] [Medline: 31888990]
56. Lerner JS, Li Y, Weber EU. The financial costs of sadness. Psychol Sci 2012 Nov 13;24(1):72-79. [doi: 10.1177/0956797612450302]
57. Garg N, Lerner JS. Sadness and consumption. J Consum Psychol 2013 Jan;23(1):106-113. [doi: 10.1016/j.jcps.2012.05.009]
58. Cohen GL, Garcia J, Purdie-Vaughns V, Apfel N, Brzustoski P. Recursive processes in self-affirmation: intervening to close the minority achievement gap. Science 2009 Apr 17;324(5925):400-403 [FREE Full text] [doi: 10.1126/science.1170769] [Medline: 19372432]
59. Logel C, Cohen GL. The role of the self in physical health: testing the effect of a values-affirmation intervention on weight loss. Psychol Sci 2012 Jan 01;23(1):53-55. [doi: 10.1177/0956797611421936] [Medline: 22157517]
60. Goyer JP, Garcia J, Purdie-Vaughns V, Binning KR, Cook JE, Reeves SL, et al. Self-affirmation facilitates minority middle schoolers' progress along college trajectories. Proc Natl Acad Sci U S A 2017 Jul 18;114(29):7594-7599 [FREE Full text] [doi: 10.1073/pnas.1617923114] [Medline: 28630338]
61. Taber JM, Klein WM, Ferrer RA, Augustson E, Patrick H. A pilot test of self-affirmations to promote smoking cessation in a national smoking cessation text messaging program. JMIR Mhealth Uhealth 2016 Jun 08;4(2):e71 [FREE Full text] [doi: 10.2196/mhealth.5635] [Medline: 27278108]
62. Augustson E, Cole-Lewis H, Sanders A, Schwarz M, Geng Y, Coa K, et al. Analysing user-reported data for enhancement of SmokefreeTXT: a national text message smoking cessation intervention. Tob Control 2017 Nov;26(6):683-689. [doi: 10.1136/tobaccocontrol-2016-052945] [Medline: 27852892]
63. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström test for nicotine dependence: a revision of the Fagerström tolerance questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
64. DiClemente CC, Prochaska JO, Fairhurst SK, Velicer WF, Velasquez MM, Rossi JS. The process of smoking cessation: an analysis of precontemplation, contemplation, and preparation stages of change. J Consult Clin Psychol 1991 Apr;59(2):295-304. [doi: 10.1037//0022-006x.59.2.295] [Medline: 2030191]
65. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988 Jun;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
66. Radler B. The Midlife in the United States (MIDUS) Series: a national longitudinal study of health and well-being. Open Health Data 2014;2(1):e3 [FREE Full text] [doi: 10.5334/ohd.ai] [Medline: 25558376]
67. Mroczek DK, Kolarz CM. The effect of age on positive and negative affect: a developmental perspective on happiness. J Pers Soc Psychol 1998 Nov;75(5):1333-1349. [doi: 10.1037//0022-3514.75.5.1333] [Medline: 9866191]

68. Scheier MF, Carver CS. Optimism, coping, and health: assessment and implications of generalized outcome expectancies. Health Psychol 1985;4(3):219-247. [doi: 10.1037//0278-6133.4.3.219] [Medline: 4029106]
69. Reed M, Aspinwall L. Self-affirmation reduces biased processing of health-risk information. Motiv Emot 1998;22:99-132. [doi: 10.1023/A:1021463221281]
70. MCQueen A, Klein WM. Experimental manipulations of self-affirmation: a systematic review. Self Identity 2006 Oct;5(4):289-354. [doi: 10.1080/15298860600805325]
71. McEwen A. The "Not-a-Puff" rule. National Centre for Smoking Cessation and Training (NCSCT). 2013. URL: http://www.ncsct.co.uk/usr/pub/not_a_puff_rule.pdf [accessed 2020-12-22]
72. National Health Service. Stop Smoking Service. 2018. URL: https://www.nhs.uk/live-well/quit-smoking/nhs-stop-smoking-services-help-you-quit/
73. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC; 2019. URL: https://www.stata.com/ [accessed 2020-12-22]
74. Mathieu E, McGeechan K, Barratt A, Herbert R. Internet-based randomized controlled trials: a systematic review. J Am Med Inform Assoc 2013 May 1;20(3):568-576 [FREE Full text] [doi: 10.1136/amiajnl-2012-001175] [Medline: 23065196]
75. Taber JM, Klein WMP, Ferrer RA, Lewis KL, Harris PR, Shepperd JA, et al. Information Avoidance Tendencies, Threat Management Resources, and Interest in Genetic Sequencing Feedback. Ann Behav Med 2015 Aug;49(4):616-621 [FREE Full text] [doi: 10.1007/s12160-014-9679-7] [Medline: 25582989]
76. Schneider S, Gadinger M, Fischer A. Does the effect go up in smoke? A randomized controlled trial of pictorial warnings on cigarette packaging. Patient Educ Couns 2012 Jan;86(1):77-83. [doi: 10.1016/j.pec.2011.03.005] [Medline: 21482060]
77. Taber JM, Klein WM, Ferrer RA, Kent EE, Harris PR. Optimism and spontaneous self-affirmation are associated with lower likelihood of cognitive impairment and greater positive affect among cancer survivors. Ann Behav Med 2016 Apr;50(2):198-209 [FREE Full text] [doi: 10.1007/s12160-015-9745-9] [Medline: 26497697]
78. Leventhal AM, Piper ME, Japuntich SJ, Baker TB, Cook JW. Anhedonia, depressed mood, and smoking cessation outcome. J Consult Clin Psychol 2014 Feb;82(1):122-129 [FREE Full text] [doi: 10.1037/a0035046] [Medline: 24219183]

Abbreviations

ANOVA: analysis of variance
OR: odds ratio
RCT: randomized controlled trial
SSAM: spontaneous self-affirmation measure

Edited by G Eysenbach; submitted 26.02.20; peer-reviewed by S Walters, J Willoughby, A Moyer; comments to author 29.06.20; revised version received 21.08.20; accepted 11.11.20; published 05.03.21.

Please cite as:
Seaman EL, Robinson CD, Crane D, Taber JM, Ferrer RA, Harris PR, Klein WMP
Association of Spontaneous and Induced Self-Affirmation With Smoking Cessation in Users of a Mobile App: Randomized Controlled Trial
J Med Internet Res 2021;23(3):e18433
URL: https://www.jmir.org/2021/3/e18433
doi:10.2196/18433
PMID:33666561

Original Paper

Designing the Optimal Digital Health Intervention for Patients' Use Before and After Elective Orthopedic Surgery: Qualitative Study

Anna Robinson[1,2], MPharm, PGDip; Robert D Slight[2,3], PhD; Andrew K Husband[1,2], PhD; Sarah P Slight[1,2], PhD

[1]School of Pharmacy, Newcastle University, Newcastle upon Tyne, United Kingdom

[2]Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom

[3]Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, United Kingdom

Corresponding Author:
Sarah P Slight, PhD
School of Pharmacy
Newcastle University
King George VI Building
Newcastle University
Newcastle upon Tyne, NE1 7RU
United Kingdom
Phone: 44 191 208 6000
Email: Sarah.slight@newcastle.ac.uk

Abstract

Background: Health behavior changes made by patients during the perioperative period can impact the outcomes and success of elective surgeries. However, there remains a limited understanding of how best to support patients during this time, particularly through the use of digital health interventions. Recognizing and understanding the potential unmet needs of elective orthopedic surgery patients is central to motivating healthier behavior change, improving recovery, and optimizing overall surgical success in the short and long term.

Objective: The aim of this study is to explore patient perspectives on technology features that would help support them to change their lifestyle behaviors during the pre- and postoperative periods, and that could potentially maintain long-term healthy lifestyles following recovery.

Methods: Semistructured interviews with pre- and postoperative elective orthopedic patients were conducted between May and June 2020 using telephone and video call–based software. Patient perspectives on the use of digital technologies to complement current surgical care and support with lifestyle behavior changes were discussed. Interviews were audio recorded and transcribed verbatim. Reflexive thematic analysis enabled the development of themes from the data, with QSR NVivo software (version 12) facilitating data management. Ethical approval was obtained from the National Health Service Health Research Authority.

Results: A total of 18 participants were interviewed. Four themes were developed from the data regarding the design and functionality of digital technologies to best support the perioperative journey. These center around an intervention's ability to incorporate interactive, user-centered features; direct a descriptive and structured recovery; enable customizable, patient-controlled settings; and deliver both general and specific surgical advice in a timely manner. Interventions that are initiated preoperatively and continued postoperatively were perceived as beneficial. Interventions designed with personalized milestones were found to better guide patients through a structured recovery. Individualized tailoring of preparatory and recovery information was desired by patients with previously high levels of physical activity before surgery. The use of personalized progression-based exercises further encouraged physical recovery; game-like rewards and incentives were regarded as motivational for making and sustaining health behavior change. In-built video calling and messaging features offered connectivity with peers and clinicians for supported care delivery.

Conclusions: Specific intervention design and functionality features can provide better, structured support for elective orthopedic patients across the entire surgical journey and beyond. This study provides much-needed evidence relating to the optimal design and timing of digital interventions for elective orthopedic surgical patients. Findings from this study suggest a desire for personalized perioperative care, in turn, supporting patients to make health behavior changes to optimize surgical success. These findings should be used to influence future co-design projects to enable the design and implementation of patient-focused, tailored, and targeted digital health technologies within modern health care settings.

(J Med Internet Res 2021;23(3):e25885) doi:10.2196/25885

KEYWORDS

digital technology; orthopedic surgery; behavior change; perioperative care; prehabilitation; qualitative research; mHealth; eHealth; mobile phone

Introduction

Background

Digital health technologies are becoming increasingly common in various industries, with medicine, surgery, and health care being no exception [1]. The use of digital health interventions is growing significantly among patients and health care providers, with recent studies reporting over 318,000 smartphone apps available to aid in health education, diagnosis, and self-management [2-4]. Despite the multitude of digital solutions available, many fail to meet patient and provider expectations—with their use and uptake hindered by ethical issues such as privacy and security of data, disease management, and communication [5]. Involving technology end users in cocreation approaches has been acknowledged as a possible strategy to design digital interventions that meet both the patient and provider needs [6,7].

In recent digital health literature, there are various interventions that have successfully supported patients in managing long-term health conditions [8] and medication adherence [9,10] and supporting positive lifestyle behavior change before and after surgery to improve postoperative outcomes [11,12]. Health behavior changes made during the perioperative period can be fundamental in determining the outcomes and success of elective surgeries. In the context of orthopedic surgery, increases in preoperative physical activity levels and smoking cessation have been associated with improved postoperative bone healing [13], wound healing [14], quicker recovery times, and reduced pain scores [15]. Physical rehabilitation after orthopedic surgery is an essential component of treatment, as it helps to improve functional outcomes and support patients to return to their daily activities [16]. There remains a limited understanding of how best to support patients during this time, particularly through the use of digital interventions.

In this context, research has focused on the orthopedic clinician's use of digital technologies [1,17], for instance, in supporting their educational development [18], guiding clinical decision support [19], managing care referrals [20], and building the patient-clinician relationship [21,22]. Recognizing and understanding the potential unmet needs of elective orthopedic surgery patients is central to motivating healthier behavior change, improving their recovery, and optimizing overall surgical success in the short and long term [23-25]. The optimal design and functionality of digital solutions to aid this cohort are yet to be recognized.

Objectives

To develop useful and effective digital technologies and strategies, it is important to first understand how patients want to be supported on their care pathway. Our patient-informed research applies qualitative investigation to explore patient perspectives and identify key technology features that they would find supportive during the pre- and postoperative periods and that could potentially maintain long-term healthy lifestyles following recovery. Specifically, our key research questions concerned the following: (1) *What* do orthopedic patients want from digital health technologies? (2) *How* do they want to use them? and (3) *When* would they be of most benefit during their elective surgical journey?

Methods

Recruitment and Sampling

The Consolidated Criteria for Reporting Qualitative Research checklist was followed for this study, according to Enhancing the QUAlity and Transparency Of health Research guidelines (Multimedia Appendix 1) [26]. Immediately before study commencement, COVID-19 restrictions were enforced across the United Kingdom. This meant that the planned face-to-face recruitment and data collection could no longer be undertaken in person at one of the largest teaching hospitals in North England. Instead, an amendment to the National Health Service (NHS) Health Research Authority (HRA) Ethics meant that participants could be recruited remotely via email and social media. All participants were emailed with an information sheet detailing the purpose and aim of this study. Participants who expressed an interest and provided written consent were enrolled in the study. There was no prior relationship established between the researcher and participants before study commencement or recruitment. Inclusion criteria were as follows: participants aged more than 18 years who were due to undergo (or had recently undergone, within the last 2 years) elective orthopedic surgery, who were medically stable and did not have an acute decline in health away from their baseline, who were able to participate in an interview, who were able to communicate in English, and who had the capacity to consent to participate in the study. Purposive sampling was used to recruit participants undergoing a variety of orthopedic surgical procedures, with mixed age ranges and sociodemographic backgrounds.

Semistructured Interviews

In-depth, semistructured interviews were conducted by 1 researcher (AR, a female doctoral researcher with experience in qualitative research) between May and June 2020 while working from home. Interviews were conducted with participants over the telephone or by using video call–based software, such as Zoom and Microsoft Teams, and all participants were offered the choice of which they would prefer. The semistructured interview topic guide was developed based on 3 pilot interviews and covered key issues identified through a systematic literature review [12], meta-ethnography [27], and narrative review [28]. These issues included participants' understanding and experiences of surgery, awareness of perioperative lifestyle behavior change, perspectives on digital

health technology use within the surgical pathways, and the optimal design of such technologies.

Data Analysis

All interviews were audio recorded and transcribed verbatim by 1 researcher (AR). All data were anonymized at the point of transcription; participants did not comment on the transcripts or provide feedback on results. Following reflexive thematic analysis processes, as defined by Braun and Clarke [29,30], each interview was transcribed and analyzed before conducting the next interview. The principle of constant comparison guided the iterative process of data collection and analysis. Two researchers (AR and AH) performed a reflexive thematic analysis to analyze the data. Close and detailed reading of the transcripts allowed the 2 researchers to familiarize themselves with the data. Initial descriptive codes were identified in a systematic manner across the data set; these were then sorted into common coding patterns, which enabled the development of analytic themes from the data. The themes were reviewed, refined, and named once coherent and distinctive. Two authors (AR and AH) performed the data analysis through discussion and, if agreement was not reached, by consensus with the wider research team (SS and RS). The postinterview field notes enhanced this reflective process. QSR NVivo software (version 12) was used to facilitate data management. The research team agreed that data saturation occurred in 18 interviews. To ensure confidentiality when using direct patient quotes within this research, nonidentifiable pseudonyms are used throughout, for example, participant 1 and participant 2.

Ethical Approval

Ethical approval was obtained from the NHS HRA and Care Research Wales (reference: 19/NE/0318), and research governance was granted by the participating NHS trust.

Results

Overview

A total of 18 participants were recruited and interviewed as part of this study (there were no refusals to partake, participant dropout, or repeat interviews). The characteristics of participants are presented in Table 1. The average age of the participants was 52 (SD 16.7) years, and the most common elective orthopedic procedure was a total hip replacement. A total of 11 interviews were conducted over the telephone and 7 were conducted using video call–based software. The average duration of the interview was 48 (SD 8.5) minutes.

Four themes were developed from the data (Figure 1) that addressed the aforementioned research questions. These themes centered on an intervention's ability to (1) incorporate interactive, user-centered features; (2) direct a descriptive and structured recovery; (3) enable customizable, patient-controlled settings; and (4) deliver both general and specific surgical advice in a timely manner. We will discuss each of these themes, in turn, illustrating patient perspectives and recommendations with direct interview quotes.

Table 1. Participant characteristics.

Participant number	Sex (M[a] or F[b])	Age (years)	Interview format	Orthopedic procedure	Pre- or postoperative	Time since surgery or time until surgery
1	F	83	Telephone	TKR[c]	Post	12 m[d]
2	M	63	Telephone	TKR	Post	6 m
3	M	63	Telephone	TKR	Post	24 m
4	F	41	Video call	THR[e]	Post	22 m
5	F	42	Video call	THR	Post	14 m
6	M	61	Telephone	THR	Post	20 m
7	M	70	Telephone	THR	Post	16 m
8	F	50	Telephone	THR	Post	8 m
9	F	69	Telephone	THR	Post	24 m
10	M	50	Video call	THR	Post	10 m
11	M	66	Telephone	TKR	Pre	2 w[f]
12	M	26	Video call	Hip FAIS[g]	Pre	4 w
13	F	62	Telephone	WL R[h]	Pre	6 w
14	M	26	Video call	ACL R[i]	Post	6 w
15	F	30	Telephone	Ankle reconstruction	Pre	1 w
16	M	24	Video call	ACL R	Post	6 m
17	M	56	Telephone	TKR	Pre	3 w
18	M	54	Video call	THR	Pre	8 w

[a]M: male.

[b]F: female.

[c]TKR: total knee replacement.

[d]m: months.

[e]THR: total hip replacement.

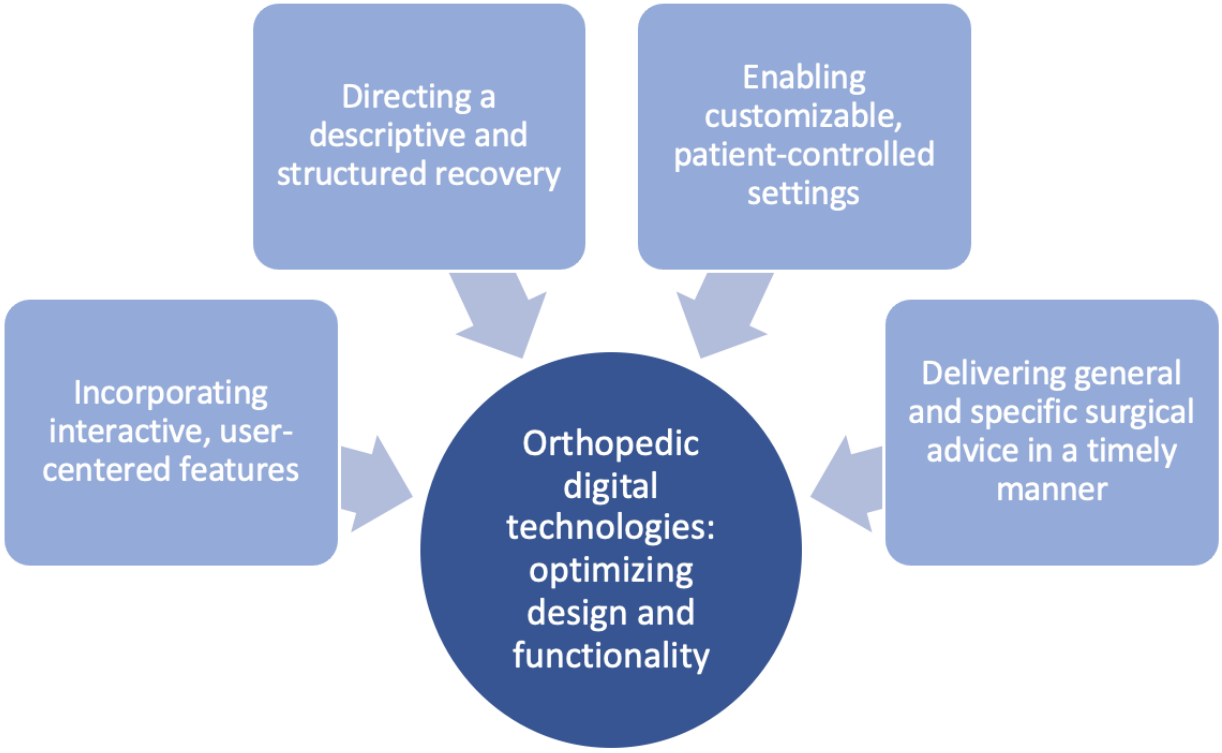
[f]w: weeks.

[g]FAIS: femoral acetabular impingement surgery.

[h]WL R: wrist ligament reconstruction.

[i]ACL R: anterior cruciate ligament reconstruction.

Figure 1. Features contributing to the design and functionality of digital technologies to best support orthopedic patients in their perioperative journey.



Theme 1: Incorporating Interactive, User-Centered Features

When considering *what* orthopedic patients wanted from digital technologies, it was important that the technology included features that centered on the needs of each user and were interactive in terms of logging and tracking and were visually instructive (video-based) and that allowed connectivity through messaging.

Logging and Tracking Recovery

Interviewees perceived numerous benefits from keeping logs during the perioperative period. In the first instance, they recognized personal benefits from "logging and tracking (their) recovery" (Participant 5) to visually "see your progress" (Participant 16) and "gauge where you are and how well you're doing" (Participant 4). This was viewed as something that would "give you the drive" to continue with the physical rehabilitation "to benefit yourself further" (Participant 9). Participants saw benefits in allowing members of the multidisciplinary team, such as surgeons and physiotherapists, to also view this information and "get an idea of when you're starting to improve" (Participant 13). Others expanded on this, seeing shared access as an opportunity to obtain further medical expert advice on "pains or swelling...and problems with the scar like any bleeding or signs of infection" (Participant 13) and find "reassurance" related to wound healing (Participant 10). Participants reflected on the accountability that can arise from shared access to logs, both as a form of intrinsic accountability to keep "helping yourself at home...to give yourself the best chance" in the recovery period (Participant 15) and as a way of "proving that you're doing what you're meant to" to their surgical team (Participant 13):

> *Things that record what you've done so you can see and say "ah, I've achieved that, I've done that"...I have the incentive to go further.* [Participant 11]

Keeping a log of wider experiences during their perioperative journey, beyond physical activity, was also considered useful. A preoperative log of "mood, sleep deprivation and pain-management" strategies was considered important for participants to "validate your (their) mental-side" in the run-up to surgery (Participant 5). Two participants reflected on personal experiences that affected their mental health during the surgical journey, where "meditation or soothing-type app" features would have supported them "through particularly tough" preoperative pain, postoperative pain, and isolation during recovery (Participant 10). Another called for integration of an interactive "diary on the app...where you could type in how you feel...if there's any problems" alongside "logging the pain and the level of pain" (Participant 13).

Within the logging features, patients described in detail about the user-centered information they wished "to be told by the technology" (Participant 5). Emphasis was placed on "the specific tracking...of any (form of) activity," rather than only walking or running; being able to "compare your times or distances" (Participant 5) through "a graph or a visual" comparative feature (Participant 11); and having real-time functionality so that you can "track your progress accurately...like keeping track of your reps and weights" without relying on retrospective data entry (Participant 16). Participants also discussed share functions when it came to their logged activities where integrated competition features, such as "leader boards with friends" (Participant 12), appeared to add an incentive to engage with physiotherapy-based recovery. Combining these with "rewards and badges" (Participant 4) for

logged activities appeared to reinforce patient motivation and engagement while keeping the technology user centered:

> *I'll think to myself "can I do it quicker, can I go further?"...which I think are all the correct messages one needs to hear when you put it in the context of general healthy living.* [Participant 11]

I Want Something to Show Me

Video features, as an interactive method of engaging with physical activity during the surgical journey, were discussed by all participants. Preoperatively, participants reflected that videos could be used to educate patients on "exercises they'll be expected to do" (Participant 5) and to build awareness around "the limitations, physically, that you will feel after the surgery" (Participant 8). Drawing on their experience, postoperative patients felt that being able to watch videos to practice rehabilitative exercises, preoperatively, would have given them "confidence and reassurance" to better engage with the recovery process from an earlier stage, optimizing "the entire recovery process, to give myself the best chance" (Participant 1). On the whole, participants felt video content to be of the highest value in the postoperative period, where patients can interact with instructional, surgery-specific rehabilitation advice by watching "user-friendly...video tutorials with people doing" the exercises (Participant 16). Participants discussed the integration of video-based postoperative "success stories" at various milestones of the recovery process, recognizing the power of video messages to help "visualize what you can achieve" (Participant 6) and "push me further with recovering" (Participant 16):

> *With each exercise there could be a video tutorial with people doing them so you can go on, click, watch the video...it could help you understand the exercise the physio(therapist) recommends...and learn how to do it properly so it's of most benefit.* [Participant 16]

Along with using videos for instructional and educational purposes, participants reflected on their "changed views" (Participant 9) of integrating video call features in digital technologies for support during the perioperative period. For many, these views were linked to and influenced by the global COVID-19 pandemic. Participants described the usefulness of video calls, accepting them as a valuable and convenient form of communication while "getting used to a new normal, a different way of doing things with technology" (Participant 13). When discussing an upcoming preassessment appointment, one participant remarked that their preference for the consultation would be a video call in comparison with a proposed telephone call:

> *I'd be more than happy with Skype to "see them" for my appointment...I think it's more personal, phone calls aren't personal...I'd much prefer to Skype now instead.* [Participant 13]

Two postoperative patients reflected on their current experiences of undergoing video-based physiotherapy sessions in light of COVID-19 measures. These participants remarked the following:

> *[the content of the sessions] doesn't differ that much from actually being in-person—you can see everything well, the resolution is good and the picture is clear, I can hear clearly.* [Participant 10]

> *everything we done the week before with the physio, we replicated on the Zoom call...everything that had been done in-person was quite easily done on the Zoom call.* [Participant 14]

Messaging Someone to Settle Your Nerves

Participants felt that the inclusion of message-based features (whether with other patients or with health care professionals) within an intervention would offer numerous benefits. Preoperative participants expressed value in communicating with other people undergoing the same surgery. This related to information-seeking needs to learn from peers, with participants discussing how they already "have looked for blogs and posts from other people going through the operation" (Participant 12) and "asked for advice...to find out what the surgery is like" (Participant 13). Coupled with searching for educational support, preoperative patients reported seeking reassurance from others before undergoing the surgery, to hear "success stories (of) people who have gone through it successfully" (Participant 17) to support their surgical decision making. This was particularly important in younger participants who wished to discuss with patients of similar ages to ask "how quick their recovery was" (Participant 15) and the level to which they retained their physical activity and functioning following surgery. One participant described their preoperative nerves without knowing "what my life is going to look like after my surgery" and considered how "having a conversation or messaging someone to settle your nerves" could help (Participant 15).

Postoperatively, participants discussed the value of sharing "experiences on a forum" (Participant 7), suggesting the integration of a patient-led "discussion area" within an app "for people who've gone through similar surgeries—whatever question it may be, they can put it on there and receive feedback from people" (Participant 16). Participants demonstrated an awareness of "mis-information or mis-interpreting the information" that may be shared (Participant 16) and acknowledged how one could become easily "disillusioned" by comparing or "judging yourself on other people's recovery" (Participant 2). Both pre- and postoperative participants considered the morale boost that can come from communicating with peers, regardless of the stage of their surgical journey:

> *It's harder when you're on your own, but when you're doing it alongside other people, having them to just be there as a point of reference or just to ask daft things to, that's much easier.* [Participant 5]

"Messaging features" (Participant 10) could also enable two-way interactivity between the patient and a member of the multidisciplinary team, where examples discussed took various forms, from real-time "live-chat boxes" (Participant 10) to "a personal account, like Facebook messenger" (Participant 13). It was important to specify response times when it came to seeking information in this manner, with some participants desiring an "instant reply from someone" (Participant 10) for emergency purposes such as "wound healing or infection"

concerns (Participant 4), whereas others considered that a "response within 24 hours...or a defined period of time" for generic questions was suitable to "fit around the (professional's) workload" (Participant 16). Participants saw value in sharing both image- and text-based messages to aid clinical decision making, such as "how is your wound healing?" (Participant 10) or "identifying any signs of infection" (Participant 3), suggesting the value of visual connectivity in this cohort. Similar feelings of reassurance were also seen when participants discussed possible interactivity with members of the surgical multidisciplinary team:

> *Even the idea of (clinicians) saying "we're here, even though it's through technology"...it gives you a bit peace of mind.* [Participant 9]

Theme 2: Directing a Descriptive and Structured Recovery Plan

Another important consideration of *what* this study wants from digital technologies was how directed, descriptive, and structured the content was. Perioperative participants expressed their desire for a digital intervention that could support them in "making the best recovery" by providing a structured and directed program with "suggestions of what you should be doing at each stage" (Participant 16). Postoperative patients discussed a "lack of direction" (Participant 16) in their current experiences following surgery, with extended periods between follow-up appointments where they lacked "the necessary, ongoing support" (Participant 2). One postoperative patient, 2 years after their total hip replacement, described gaps where "I was just winging it, really" in relation to recommended physiotherapy exercises:

> *there was no kind of updates with stuff when I was at home.* [Participant 9]

Participants reported knowing "within each stage of recovery, you should be pushing a little bit more" (Participant 15) but felt unsupported to do this. This view was especially apparent in previously physically active patients and those of a younger demographic who wanted to be challenged further to restore "functionality in the joint after surgery" (Participant 12):

> *all I was after was some indication of what to do to safely push on...having some indication of "this is what you need to do in this week, then move onto this"...I wanted something to show me.* [Participant 4]

Recommendations to provide this structured recovery program stemmed around designing "milestones...in terms of where you could expect to be after Week 1, Week 2," with the inclusion of "physiotherapy messages" (Participant 12) and "general healthy living messages" (Participant 11). Tailoring the intervention to support a structured recovery would mean starting with "simple exercises to start the recovery and build on from there" (Participant 16). Participants described the integration of gamification features and "progression-based exercises" throughout the recovery where, over time, the program recommended "trickier exercises...working towards that final goal of being recovered" (Participant 16). Both pre- and postoperative participants viewed the capability of setting "targets and goals to work towards" as an important feature of creating a structured and directed recovery program (Participant 4). Combining goal setting with gamification features to break "(rehabilitation) down into small chunks at the start, advancing through each level" (Participant 15) and real-time messages of support such as "well done, you've completed this level, next it's..." (Participant 4) were deemed motivational in giving "more people focus for what to achieve after the surgery" (Participant 5). Having a directed rehabilitation structure with set *milestones* to unlock over time also allowed participants "to feel some independence that it's up to you to advance through the levels or reach a certain target, but with the comfort of knowing it's still safe, you're not pushing too hard" (Participant 12). The incorporation of safety-netting features to recover at *a safe speed* also provided reassurance for preoperative patients that they will not be pushed to "do too much too soon" (Participant 12) and compromise their outcomes following surgery.

Theme 3: Enabling Customizable, Patient-Controlled Settings

When it came to addressing our research question of *how* patients wished to use these technologies, the benefits of having built-in, customizable, and "patient-controlled features" to enable elements of control were widely discussed (Participant 4). This ranged from wanting to "set myself my profile, choose my name..." (Participant 14) to having the ability to "build your own workout" (Participant 16) and "preference certain exercises to make it individualized to each person" (Participant 12). Interviewees perceived that customizable functionality would encourage greater engagement and a sense of accountability, meaning they better "connect with the (recovery) process" (Participant 4). One participant referenced the layout features of an app they were currently using, explaining how it was possible to "toggle the home-screen settings" to make it more personal (Participant 12):

> *It's going to need a personal approach—but if you were able to toggle certain settings to make it individualized to each person, then you'll get more successful outcomes with it and impact different people in different ways.* [Participant 12]

Accompanying the ability to customize aspects of physical recovery, participants also recognized benefits in preferring features relating to the mental and motivational postsurgical journey. Choosing a "more personal reminder" (Participant 7) approach to notifications was deemed constructive and supportive, with encouraging messages of "have you done your physio yet?" rather than "automated "do your physio" notifications" (Participant 12).

Having the capacity to tailor preparatory and recovery information to individual participants was widely discussed—in particular, by those who described high levels of physical activity before surgery and a wish to continue this postoperatively:

> *it completely depends on who you are as an individual and what you want from it (surgery)* [Participant 4]

Being able to "advance at a pace suitable for you" (Participant 12) during recovery was deemed imperative to restore previous

"functionality of the joint" (Participant 6) and, in the process, meet individual postoperative expectations. From their experiences, some viewed rehabilitation exercises as "rather pedestrian" (Participant 6) and that "the whole process, the whole support...was geared around older and less mobile people" (Participant 4). It appeared that exercises were not designed with a younger or more active patient in mind. When it came to using technology to manage this, participants expressed desires to be able to "choose your own difficulty...to make the recovery challenging enough" (Participant 12):

> *(recommendations) should be determined by how active you already are...it's no good telling me "walk 1 mile" when I'm used to walking 20! It's the same for someone perhaps less active when they can't functionally do it.* [Participant 4]

Theme 4: Delivering General and Specific Surgical Advice in a Timely Manner

Addressing our research question around *when* digital technologies would be of most benefit, the timing (initiation point) of the intervention appeared crucial. It was discussed that technology should be initiated to meet both the pre- and postoperative information-seeking needs of participants. Specifically, preoperative interviewees wished to have explicit "sections for before surgery" (Participant 11) to seek information about the surgical procedure, to understand the best way to prepare, and to familiarize themselves with the upcoming process of recovery so as to be "already in that mindset...to the idea of the time and energy we need to invest in order to fully recover" (Participant 17). On reflection, some postoperative patients felt that their recovery would have benefited from knowing this information in advance. "Staggering the information" was also considered important, with ideas of drip feeding and building up advice in the preoperative period so that postoperatively, they would be better prepared (Participant 10):

> *I was ready for the off, straight away...I had it in my mind that that's what I needed to do...you don't want to be waiting 'til you're post (-operative) to hear those things.* [Participant 5]

Participants felt that the initiation of digital interventions should also be arranged with a sense of *generalizability* between surgical procedures so that patients undergoing any form of elective orthopedic surgery may find the preoperative information beneficial. Participants described the need for "a generic advice" hub (Participant 15) for all orthopedic patients to use, with "different tabs for different surgeries" so that patients could find surgery-specific information if they wanted (Participant 8). Two participants discussed the feasibility of having one "centralized database" (Participant 12) of exercises, breaking "the exercises down to different body parts," and being able to easily find those that they could do to aid their recovery (Participant 16). In addition, interviewees called for holistic "general health and recovery" sections, integrating "positive health advice" that would be useful to hear throughout the perioperative process of any surgery (Participant 6). This included preoperative advice on preparation for surgery and "building muscle strength beforehand" (Participant 15), reassurance on postoperative physical rehabilitation, and "short- and long-term messages" around overall healthy living (Participant 11):

> *There are generic exercises that would be recommended for most joint surgeries, just to build up the muscle strength again...(and) if you had an app where you could select "hip replacement" and it provided you with "this is what exercises you should do"...it could give you more specific information.* [Participant 4]

Discussion

Principal Findings

This patient-informed study underlines the importance of obtaining orthopedic surgical patients' perspectives in relation to the design and functionality of digital technologies to best support their recovery. By collecting both pre- and postoperative patient perspectives, we were able to clearly identify specific features and functionalities that appear to be the most desired and of most benefit in supporting this surgical cohort across the whole perioperative pathway. We addressed 3 research areas: *what* patients wanted from digital technologies, *how* they wanted to use them, and *when* their use would be of most benefit.

A consistent finding across interviews was that participants saw value in having a digital intervention to direct them through a structured plan to achieve a *successful recovery*. In terms of technology design, both prescriptive and descriptive contents were desired, where participants called for regular digital milestones to guide them and measure their journey toward recovery. Previous studies have demonstrated the benefits of continuous measurement within the recovery process following cardiac [31] and neurological surgery [32]. This feature should be considered for orthopedic interventions, where quantifying the progress can motivate patients to take active roles in their recovery [33]. Mehta et al [34] aligned this idea with reports of positive reinforcement by setting and meeting individual recovery goals following hip arthroplasty. Goal setting is a well-recognized behavior change technique that supports self-regulation skills in the change process [35,36]. In previous orthopedic studies, digital goal setting facilitated personal fulfillment and provided patients with a sense of control and accomplishment during the perioperative period [37,38]. Combining goal setting with performance feedback and the review of goals (akin to milestones within the recovery journey) has been associated with both short- and long-term intervention effectiveness [39,40]. Personalized and tailored feedback on these goals could be acknowledged as relevant and actionable, as opposed to generic advice [41]. By integrating digital strategies to help define goals within recovery, orthopedic patients may feel better supported and motivated to engage in health behavior change.

Participants valued the integration of video-based features in digital interventions, whether as a visual aid for rehabilitative exercises or to facilitate remote telemedicine consultations. Our findings support the growing popularity of video-based consultations reported in other areas of global health and social care [42-44], with participants reporting feelings of

connectedness, empowerment, and reassurance through image- and video-based sharing [45-47]. The incorporation of video call features within digital health technology is gaining attention, particularly as a consequence of the global COVID-19 pandemic [43,48]. It appeared that the more prominent use of video call features, both in participants' work and social lives, has led to greater acceptance and adoption of their use within the world of health care [48].

Another promising strategy of digital intervention design, *gamification*, has also been linked to increased user engagement [49,50]. In this study, participants' suggestions to incorporate leaderboards and collect rewards during the postoperative recovery process echo recent findings from adult and pediatric patients undergoing orthopedic, dental, and ophthalmic surgeries [51,52] and those concerning wider eHealth design [40,53]. The use of game-like rewards and incentives has been shown to motivate and sustain health habits over time [54,55]. In wider public health initiatives, incentive-based health apps and activity-tracking programs have been associated with positive physical activity behavior change in Canada [56] and the United Kingdom [57,58]. Other successful digital health interventions have incorporated gamification features, promoting intrinsic and extrinsic motivators [59-61]. Similarities can also be recognized between these findings and wider work on persuasive systems design in relation to shaping health behaviors during the perioperative period [62-66]. Oinas-Kukkonen and Harjumaa [62] proposed that persuasion principles (including praise and rewards) should be considered as requirements in software design.

This study contributes further evidence to support gaps in the literature, which relate to the timing of intervention use, including the initiation and continuation points of intervention use. This gap has also been acknowledged in recent systematic reviews and research by Jansson et al [16], Mirkovic et al [67], and our research team [12,27]. Interventions that are initiated preoperatively and continued postoperatively were perceived as beneficial. Captivating the preoperative patient mindset and making use of the surgical teachable moment appears to be significant in encouraging perioperative behavior change and optimizing postoperative outcomes [68]. Being granted a sense of control and responsibility over their recovery by initiating and using interventions preoperatively were valued by participants. Before surgery, interviewees described the desire to customize their technology and its content to best suit their needs, thereby encouraging better engagement with the upcoming recovery process. The individualization of care pathways has been discussed in medical and surgical literature [69,70]; however, our study also highlights the importance of individualization of the *technologies* to support care delivery. Technologies that incorporated customizable features, which the patient could control and toggle according to their personal preferences, were considered another motivator for successful recovery. Participant autonomy has been shown to positively impact motivation levels and user experience, thereby improving patient care experiences [71-73]. Technology-enabled, preference-based care has improved patient and health care professional outcomes [72-74]. Technology creators may consider implementing customizable features to grant patients autonomy over aspects of their recovery [67,75].

All participants in this study discussed the impact of the global COVID-19 pandemic on the UK NHS. At the point of interview, 3 participants were undergoing technology-enabled follow-up appointments with their physiotherapist and 2 had used video call–based software to conduct their preoperative assessments with members of the surgical multidisciplinary team. Participants' views echoed those discussed in this study on *digitally engaged patients* and recognized the multitude of ways in which technologies can be embedded within the NHS to transform surgical patient support throughout the entire perioperative journey [48]. Interactive health technologies have been credited as transformers of health care by supporting engaged self-care and promoting positive health behaviors [76]. The global pandemic has presented a unique opportunity for the creative delivery of health care. It is important that this momentum gained to adopt and use digital technologies is not lost, with the focus being continued provision of innovative surgical patient care, monitoring, and follow-up spanning the whole perioperative period [77].

Limitations

We acknowledge that there are some limitations with this study. The intended method of in-person data collection was impacted by the COVID-19 pandemic. Although virtual call–based software enabled the replication of face-to-face interviews (ie, responding to verbal and nonverbal cues and building rapport) [78,79], there are some disadvantages to this interview technique that may have impacted our study. Established familiarity and participant comfort of use may have resulted in the higher number of interviews conducted over the telephone. Despite this, video calls enabled a unique snapshot into life of a patient recovering at home during the crisis and provided a fuller picture with more context than a telephone call may have done [80]. Participants currently experiencing remote consultations with members of the surgical team offered timely insights into this study. As a result of the COVID-19 pandemic, many elective orthopedic surgeries were canceled, which meant that fewer preoperative participants could be recruited and interviewed in comparison with postoperative participants. This study predominantly focused on a small sample of patients in Northern England, and as a result, the experiences shared by participants may not be representative of all care pathways across the United Kingdom. This study also focused solely on the perspectives of elective orthopedic surgical patients, and thus, the results may not be generalizable to other elective surgical specialties or acute surgeries.

Conclusions

The results of this study have important implications for the design, functionality, application, and use of digital technologies for patients undergoing elective orthopedic surgery. By integrating digital goal-setting strategies within their recovery, patients feel better supported and motivated to engage in health behavior change to optimize surgical outcomes. The use of game-like rewards and incentives has been seen to motivate and sustain positive health habits over time. The integration of video features was acknowledged as an interactive method of engaging

with physical activity during recovery and is regarded as a more personal strategy to enable follow-up consultations. This study contributes to the limited amount of existing digital health literature in this patient cohort and provides much-needed evidence relating to the optimal timing of digital interventions for elective orthopedic surgical patients. These findings should be employed in future codesign projects to enable the design and implementation of patient-focused, tailored, and targeted digital health technologies within modern health care settings.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Consolidated Criteria for Reporting Qualitative Research: 32-item checklist.
[DOCX File , 23 KB - jmir_v23i3e25885_app1.docx]

References

1. Al-Obaidi B, Babu S, Al-Hadithy N. Smartphones in orthopaedics: a decade of innovation. Br J Hosp Med (Lond) 2019 May 02;80(5):274-277. [doi: 10.12968/hmed.2019.80.5.274] [Medline: 31059338]
2. The growing value of digital health in the United Kingdom. The IQVIA Institute. 2017 Nov 01. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health-in-the-united-kingdom [accessed 2021-02-02]
3. Armstrong KA, Coyte PC, Brown M, Beber B, Semple JL. Effect of home monitoring via mobile app on the number of in-person visits following ambulatory surgery: a randomized clinical trial. JAMA Surg 2017 Mar 22;152(7):622-627. [doi: 10.1001/jamasurg.2017.0111] [Medline: 28329223]
4. Schnall R, Mosley JP, Iribarren SJ, Bakken S, Carballo-Diéguez A, Brown IW. Comparison of a user-centered design, self-management app to existing mhealth apps for persons living with HIV. JMIR Mhealth Uhealth 2015;3(3):e91 [FREE Full text] [doi: 10.2196/mhealth.4882] [Medline: 26385783]
5. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. Acad Emerg Med 2015 Jun;22(6):754-756 [FREE Full text] [doi: 10.1111/acem.12692] [Medline: 25997375]
6. Brennan L, Kessie T, Caulfield B. Patient experiences of rehabilitation and the potential for an mHealth system with biofeedback after breast cancer surgery: qualitative study. JMIR Mhealth Uhealth 2020 Jul 29;8(7):e19721 [FREE Full text] [doi: 10.2196/19721] [Medline: 32687476]
7. Burns D, Razmjou H, Shaw J, Richards R, McLachlin S, Hardisty M, et al. Adherence tracking with smart watches for shoulder physiotherapy in rotator cuff pathology: protocol for a longitudinal cohort study. JMIR Res Protoc 2020 Jul 05;9(7):e17841 [FREE Full text] [doi: 10.2196/17841] [Medline: 32623366]
8. Apolinário-Hagen J, Menzel M, Hennemann S, Salewski C. Acceptance of mobile health apps for disease management among people with multiple sclerosis: web-based survey study. JMIR Form Res 2018 Dec 12;2(2):e11977 [FREE Full text] [doi: 10.2196/11977] [Medline: 30684408]
9. Santo K, Singleton A, Rogers K, Thiagalingam A, Chalmers J, Chow CK, et al. Medication reminder applications to improve adherence in coronary heart disease: a randomised clinical trial. Heart 2019 Feb 27;105(4):323-329. [doi: 10.1136/heartjnl-2018-313479] [Medline: 30150326]
10. Sua YS, Jiang Y, Thompson DR, Wang W. Effectiveness of mobile phone-based self-management interventions for medication adherence and change in blood pressure in patients with coronary heart disease: a systematic review and meta-analysis. Eur J Cardiovasc Nurs 2020 Mar 19;19(3):192-200. [doi: 10.1177/1474515119895678] [Medline: 31856596]
11. McDonald S, Yates D, Durrand JW, Kothmann E, Sniehotta FF, Habgood A, et al. Exploring patient attitudes to behaviour change before surgery to reduce peri-operative risk: preferences for short- vs. long-term behaviour change. Anaesthesia 2019 Dec 21;74(12):1580-1588 [FREE Full text] [doi: 10.1111/anae.14826] [Medline: 31637700]
12. Robinson AH, Husband AK, Slight RD, Slight SP. Effectiveness of digital technologies to support surgical patients in changing their health behaviours: a systematic review. BJS Open 2020 Dec 28:1-4 [FREE Full text] [doi: 10.1093/bjsopen/zraa009]
13. Truntzer J, Vopat B, Feldstein M, Matityahu A. Smoking cessation and bone healing: optimal cessation timing. Eur J Orthop Surg Traumatol 2015 Feb 31;25(2):211-215. [doi: 10.1007/s00590-014-1488-y] [Medline: 24879610]
14. Sørensen LT. Wound healing and infection in surgery. The clinical impact of smoking and smoking cessation: a systematic review and meta-analysis. Arch Surg 2012 Apr 01;147(4):373-383. [doi: 10.1001/archsurg.2012.5] [Medline: 22508785]
15. Luna IE, Kehlet H, Wede HR, Hoevsgaard SJ, Aasvang EK. Objectively measured early physical activity after total hip or knee arthroplasty. J Clin Monit Comput 2019 Jun 23;33(3):509-522. [doi: 10.1007/s10877-018-0185-5] [Medline: 30039461]
16. Jansson MM, Rantala A, Miettunen J, Puhto A, Pikkarainen M. The effects and safety of telerehabilitation in patients with lower-limb joint replacement: a systematic review and narrative synthesis. J Telemed Telecare 2020 Apr 21:1357633X20917868. [doi: 10.1177/1357633X20917868] [Medline: 32316837]

17. Popat R, Mohan AT, Branford OA. Current uses of smartphones and apps in orthopaedic surgery. Br J Hosp Med (Lond) 2013 Dec;74(12):672-676. [doi: 10.12968/hmed.2013.74.12.672] [Medline: 24326713]
18. Chen EA, Ellahie A, Barsi J. Smartphone applications in orthopaedic surgery: a review of the literature and application analysis. Current Orthopaedic Practice 2019;30:230. [doi: 10.1097/bco.0000000000000745]
19. Watson HA, Tribe RM, Shennan AH. The role of medical smartphone apps in clinical decision-support: a literature review. Artif Intell Med 2019 Sep;100:101707-101704. [doi: 10.1016/j.artmed.2019.101707] [Medline: 31607347]
20. Morkel RW, Mann TN, Du Preez G, Du Toit J. Orthopaedic referrals using a smartphone app: uptake, response times and outcome. S Afr Med J 2019 Oct 31;109(11):859. [doi: 10.7196/samj.2019.v109i11.13986]
21. De La Cruz Monroy MFI, Mosahebi A. The use of smartphone applications (Apps) for enhancing communication with surgical patients: a systematic review of the literature. Surg Innov 2019 Apr;26(2):244-259. [doi: 10.1177/1553350618819517] [Medline: 30602332]
22. Sharif FS, Rahman A, Tonner E, Ahmed H, Haq I, Abbass R, et al. Can technology optimise the pre-operative pathway for elective hip and knee replacement surgery: a qualitative study. Perioper Med (Lond) 2020 Nov 16;9(1):33 [FREE Full text] [doi: 10.1186/s13741-020-00166-0] [Medline: 33292556]
23. Tang M, Li Z, Dai Y, Li X. What kind of a mobile health app do patients truly want? A pilot study among ambulatory surgery patients. Patient Prefer Adherence 2019;13:2039-2046 [FREE Full text] [doi: 10.2147/PPA.S220207] [Medline: 31824139]
24. Datillo JR, Gittings DJ, Sloan M, Hardaker WM, Deasey MJ, Sheth NP. "Is There An App For That?" Orthopaedic patient preferences for a smartphone application. Appl Clin Inform 2017 Aug 16;8(3):832-844 [FREE Full text] [doi: 10.4338/ACI-2017-04-RA-0058] [Medline: 28885637]
25. Michard F, Gan T, Kehlet H. Digital innovations and emerging technologies for enhanced recovery programmes. Br J Anaesth 2017 Jul 01;119(1):31-39 [FREE Full text] [doi: 10.1093/bja/aex140] [Medline: 28605474]
26. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
27. Robinson A, Oksuz U, Slight R, Slight S, Husband A. Digital and Mobile Technologies to Promote Physical Health Behavior Change and Provide Psychological Support for Patients Undergoing Elective Surgery: Meta-Ethnography and Systematic Review. JMIR Mhealth Uhealth 2020 Dec 01;8(12):e19237 [FREE Full text] [doi: 10.2196/19237] [Medline: 33258787]
28. Robinson A, Husband AK, Slight RD, Slight SP. Digital support for patients undergoing bariatric surgery: narrative review of the roles and challenges of online forums. JMIR Perioper Med 2020 Jul 15;3(2):e17230 [FREE Full text] [doi: 10.2196/17230] [Medline: 33393926]
29. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
30. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
31. Cook DJ, Thompson JE, Prinsen SK, Dearani JA, Deschamps C. Functional recovery in the elderly after major surgery: assessment of mobility recovery using wireless technology. Ann Thorac Surg 2013 Sep;96(3):1057-1061. [doi: 10.1016/j.athoracsur.2013.05.092] [Medline: 23992697]
32. Appelboom G, Taylor BE, Bruce E, Bassile CC, Malakidis C, Yang A, et al. Mobile phone-connected wearable motion sensors to assess postoperative mobilization. JMIR Mhealth Uhealth 2015 Jul 28;3(3):e78 [FREE Full text] [doi: 10.2196/mhealth.3785] [Medline: 26220691]
33. Lyman S, Hidaka C, Fields K, Islam W, Mayman D. Monitoring patient recovery after THA or TKA using mobile technology. HSS J 2020 Dec;16(Suppl 2):358-365. [doi: 10.1007/s11420-019-09746-3] [Medline: 33380968]
34. Mehta N, Steiner C, Fields KG, Nawabi DH, Lyman SL. Using mobile tracking technology to visualize the trajectory of recovery after hip arthroscopy: a case report. HSS J 2017 Jul;13(2):194-200 [FREE Full text] [doi: 10.1007/s11420-017-9544-x] [Medline: 28690471]
35. Michie S, Whittington C, Hamoudi Z, Zarnani F, Tober G, West R. Identification of behaviour change techniques to reduce excessive alcohol consumption. Addiction 2012 Aug;107(8):1431-1440. [doi: 10.1111/j.1360-0443.2012.03845.x] [Medline: 22340523]
36. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
37. Argent R, Slevin P, Bevilacqua A, Neligan M, Daly A, Caulfield B. Wearable sensor-based exercise biofeedback for orthopaedic rehabilitation: a mixed methods user evaluation of a prototype system. Sensors (Basel) 2019 Jan 21;19(2):1-432 [FREE Full text] [doi: 10.3390/s19020432] [Medline: 30669657]
38. Lee M, Suh D, Son J, Kim J, Eun S, Yoon B. Patient perspectives on virtual reality-based rehabilitation after knee surgery: importance of level of difficulty. J Rehabil Res Dev 2016;53(2):239-252 [FREE Full text] [doi: 10.1682/JRRD.2014.07.0164] [Medline: 27149529]
39. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

40. Jessen S, Mirkovic J, Ruland CM. Creating gameful design in mhealth: a participatory co-design approach. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11579 [FREE Full text] [doi: 10.2196/11579] [Medline: 30552080]
41. Amann J, Fiordelli M, Scheel-Sailer A, Brach M, Rubinelli S. Opportunities and challenges of a self-management app to support people with spinal cord injury in the prevention of pressure injuries: qualitative study. JMIR Mhealth Uhealth 2020 Dec 09;8(12):e22452 [FREE Full text] [doi: 10.2196/22452] [Medline: 33295876]
42. Barsom EZ, Jansen M, Tanis PJ, van de Ven AWH, Blussé van Oud-Alblas M, Buskens CJ, et al. Video consultation during follow up care: effect on quality of care and patient- and provider attitude in patients with colorectal cancer. Surg Endosc 2021 Mar;35(3):1278-1287. [doi: 10.1007/s00464-020-07499-3] [Medline: 32198552]
43. Li HL, Chan YC, Huang JX, Cheng SW. Pilot study using telemedicine video consultation for vascular patients' care during the Covid-19 period. Ann Vasc Surg 2020 Oct;68:76-82 [FREE Full text] [doi: 10.1016/j.avsg.2020.06.023] [Medline: 32562832]
44. Rabinovich L, Molton JS, Ooi WT, Paton NI, Batra S, Yoong J. Perceptions and acceptability of digital interventions among tuberculosis patients in cambodia: qualitative study of video-based directly observed therapy. J Med Internet Res 2020 Jul 27;22(7):e16856 [FREE Full text] [doi: 10.2196/16856] [Medline: 32716309]
45. Hasselberg M, Beer N, Blom L, Wallis LA, Laflamme L. Image-based medical expert teleconsultation in acute care of injuries. A systematic review of effects on information accuracy, diagnostic validity, clinical outcome, and user satisfaction. PLoS One 2014;9(6):e98539 [FREE Full text] [doi: 10.1371/journal.pone.0098539] [Medline: 24887257]
46. Vlahu-Gjorgievska E, Nagapuri S, Win KT. Tele-monitoring technology as a tool for monitoring and management of patients with congestive heart failure. Acad J Interdiscip Stud 2019 Feb 12;23:e1. [doi: 10.3127/ajis.v23i0.1816]
47. Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. J Med Internet Res 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
48. Sahu D, Rathod V. Telerehabilitation during COVID 19: Shoulder rehabilitation: letter to the editor regarding Menendez et al: "Orthopedic surgery post COVID-19: an opportunity for innovation and transformation". J Shoulder Elbow Surg 2020 Oct;29(10):1957-1958 [FREE Full text] [doi: 10.1016/j.jse.2020.07.001] [Medline: 32659460]
49. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
50. Cotton V, Patel MS. Gamification use and design in popular health and fitness mobile applications. Am J Health Promot 2019 Mar;33(3):448-451 [FREE Full text] [doi: 10.1177/0890117118790394] [Medline: 30049225]
51. Kontadakis G, Chasiouras D, Proimaki D, Halkiadakis M, Fyntikaki M, Mania K. Gamified platform for rehabilitation after total knee replacement surgery employing low cost and portable inertial measurement sensor node. Multimed Tools Appl 2018 Sep 12;79(5-6):3161-3188. [doi: 10.1007/s11042-018-6572-6]
52. Ryu JH, Park JW, Nahm FS, Jeon YT, Oh AY, Lee HJ, et al. The effect of gamification through a virtual reality on preoperative anxiety in pediatric patients undergoing general anesthesia: a prospective, randomized, and controlled trial. J Clin Med 2018 Sep 17;7(9):284 [FREE Full text] [doi: 10.3390/jcm7090284] [Medline: 30227602]
53. Jessen S, Mirkovic J, Nes LS. Mystrengths, a strengths-focused mobile health tool: participatory design and development. JMIR Form Res 2020 Jul 24;4(7):e18049 [FREE Full text] [doi: 10.2196/18049] [Medline: 32706651]
54. King D, Greaves F, Exeter C, Darzi A. 'Gamification': influencing health behaviours with games. J R Soc Med 2013 Mar;106(3):76-78 [FREE Full text] [doi: 10.1177/0141076813480996] [Medline: 23481424]
55. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014 Aug 04;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]
56. Mitchell M, White L, Lau E, Leahey T, Adams MA, Faulkner G. Evaluating the carrot rewards app, a population-level incentive-based intervention promoting step counts across two canadian provinces: quasi-experimental study. JMIR Mhealth Uhealth 2018 Sep 20;6(9):e178 [FREE Full text] [doi: 10.2196/mhealth.9912] [Medline: 30148712]
57. Elliott M, Eck F, Khmelev E, Derlyatka A, Fomenko O. Physical activity behavior change driven by engagement with an incentive-based app: evaluating the impact of sweatcoin. JMIR Mhealth Uhealth 2019 Jul 08;7(7):e12445 [FREE Full text] [doi: 10.2196/12445] [Medline: 31287064]
58. Hajat C, Hasan A, Subel S, Noach A. The impact of short-term incentives on physical activity in a UK behavioural incentives programme. NPJ Digit Med 2019 Sep 16;2(1):91 [FREE Full text] [doi: 10.1038/s41746-019-0164-3] [Medline: 31531396]
59. Fox LG, Barnes D. Boosterbuddy: using gamification as a compensatory strategy for motivational deficits. Psychiatr Serv 2016 Jan;67(1):141-142. [doi: 10.1176/appi.ps.661101] [Medline: 26725495]
60. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized controlled trial of superbetter, a smartphone-based/internet-based self-help tool to reduce depressive symptoms. Games Health J 2015 Jun;4(3):235-246. [doi: 10.1089/g4h.2014.0046] [Medline: 26182069]
61. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
62. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24:e1. [doi: 10.17705/1cais.02428]

63. Parada F, Martínez V, Espinosa HD, Bauer S, Moessner M. Using persuasive systems design model to evaluate. Telemed J E Health 2020 Feb;26(2):251-254 [FREE Full text] [doi: 10.1089/tmj.2018.0272] [Medline: 30668227]
64. Hassan L, Xi N, Gurkan B, Koivisto J, Hamari J. Gameful self-regulation: a study on how gamified self-tracking features evoke gameful experiences. In: Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020 Presented at: 53rd Hawaii International Conference on System Sciences; Jan 7-10, 2020; Maui, Hawaii p. 1-10 URL: https://tinyurl.com/3yprrpa2 [doi: 10.24251/hicss.2020.138]
65. Asbjørnsen RA, Wentzel J, Smedsrød ML, Hjelmesæth J, Clark MM, Solberg Nes L, et al. Identifying persuasive design principles and behavior change techniques supporting end user values and needs in ehealth interventions for long-term weight loss maintenance: qualitative study. J Med Internet Res 2020 Nov 30;22(11):e22598 [FREE Full text] [doi: 10.2196/22598] [Medline: 33252347]
66. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
67. Mirkovic J, Kristjansdottir OB, Stenberg U, Krogseth T, Stange KC, Ruland CM. Patient insights into the design of technology to support a strengths-based approach to health care. JMIR Res Protoc 2016 Aug 24;5(3):e175 [FREE Full text] [doi: 10.2196/resprot.5906] [Medline: 27558951]
68. Robinson A, Slight R, Husband A, Slight S. The value of teachable moments in surgical patient care and the supportive role of digital technologies. Perioper Med (Lond) 2020;9:2 [FREE Full text] [doi: 10.1186/s13741-019-0133-z] [Medline: 32042404]
69. Grocott MPW, Edwards M, Mythen MG, Aronson S. Peri-operative care pathways: re-engineering care to achieve the 'triple aim'. Anaesthesia 2019 Jan 02;74 Suppl 1:90-99 [FREE Full text] [doi: 10.1111/anae.14513] [Medline: 30604413]
70. Levy N, Grocott MPW, Carli F. Patient optimisation before surgery: a clear and present challenge in peri-operative care. Anaesthesia 2019 Jan;74 Suppl 1:3-6 [FREE Full text] [doi: 10.1111/anae.14502] [Medline: 30604412]
71. Güldenpfennig F, Mayer P, Panek P, Fitzpatrick G. An autonomy-perspective on the design of assistive technology experiences of people with multiple sclerosis. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; Glasgow, Scotland Uk. 2019 May Presented at: CHI: Conference on Human Factors in Computing Systems; May 2019; United Kingdom p. e1. [doi: 10.1145/3290605.3300357]
72. Calvo RA, Peters D, Johnson D, Rogers Y. Autonomy in technology design. In: Proceedings of CHI'14 Extended Abstracts on Human Factors in Computing Systems. 2014 Apr Presented at: CHI'14 Extended Abstracts on Human Factors in Computing Systems; April 2014; Toronto Ontario Canada p. e1 URL: https://dl.acm.org/doi/proceedings/10.1145/2559206 [doi: 10.1145/2559206]
73. Burton E, Clayville K, Goldsmith J, Mattei N, editors. The heart of the matter: patient autonomy as a model for the wellbeing of technology users. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. 2019 Jan Presented at: AIES '19: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society; January 2019; Honolulu HI USA URL: https://dl.acm.org/doi/proceedings/10.1145/3306618 [doi: 10.1145/3306618.3314254]
74. Ruland CM. Handheld technology to improve patient care: evaluating a support system for preference-based care planning at the bedside. J Am Med Inform Assoc 2002;9(2):192-201 [FREE Full text] [doi: 10.1197/jamia.m0891] [Medline: 11861634]
75. Solem IKL, Varsi C, Eide H, Kristjansdottir OB, Mirkovic J, Børøsund E, et al. Patients' needs and requirements for ehealth pain management interventions: qualitative study. J Med Internet Res 2019 Apr 01;21(4):e13205 [FREE Full text] [doi: 10.2196/13205] [Medline: 30877780]
76. Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, et al. Back to the future: achieving health equity through health informatics and digital health. JMIR Mhealth Uhealth 2020 Jan 14;8(1):e14512 [FREE Full text] [doi: 10.2196/14512] [Medline: 31934874]
77. Menendez ME, Jawa A, Haas DA, Warner JJP, Codman Shoulder Society. Orthopedic surgery post COVID-19: an opportunity for innovation and transformation. J Shoulder Elbow Surg 2020 Jun;29(6):1083-1086 [FREE Full text] [doi: 10.1016/j.jse.2020.03.024] [Medline: 32312643]
78. Archibald MM, Ambagtsheer RC, Casey MG, Lawless M. Using zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants. Int J Qual Methods 2019 Sep 11. [doi: 10.1177/1609406919874596]
79. Lobe B, Morgan D, Hoffman KA. Qualitative data collection in an era of social distancing. Int J Qual Methods 2020 Jul 07. [doi: 10.1177/1609406920937875]
80. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. Br Med J 2020 Mar 12;368:e1 [FREE Full text] [doi: 10.1136/bmj.m998] [Medline: 32165352]

Abbreviations

HRA: Health Research Authority
NHS: National Health Service

Edited by M Focsa; submitted 19.11.20; peer-reviewed by A Pattison, D Szinay, A van der Horst; comments to author 11.01.21; revised version received 25.01.21; accepted 26.01.21; published 08.03.21.

Please cite as:
Robinson A, Slight RD, Husband AK, Slight SP
Designing the Optimal Digital Health Intervention for Patients' Use Before and After Elective Orthopedic Surgery: Qualitative Study
J Med Internet Res 2021;23(3):e25885
URL: https://www.jmir.org/2021/3/e25885
doi:10.2196/25885
PMID:33683208

Original Paper

Online Mindfulness-Based Cognitive Behavioral Therapy Intervention for Youth With Major Depressive Disorders: Randomized Controlled Trial

Paul Ritvo[1,2,3], PhD; Yuliya Knyahnytska[3,4,5], MD, PhD; Meysam Pirbaglou[1], PhD; Wei Wang[4,5], PhD; George Tomlinson[6,7], PhD; Haoyu Zhao[3,4,5], PhD; Renee Linklater[4,8], PhD; Shari Bai[4,5], MD; Megan Kirk[1], MA; Joel Katz[1,9], PhD; Lillian Harber[4], BA; Zafiris Daskalakis[3,5,10,11], MD, PhD

[1]School of Kinesiology and Health Sciences, York University, Toronto, ON, Canada

[2]Department of Psychology, York University, Toronto, ON, Canada

[3]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[4]Centre for Addiction and Mental Health, Toronto, ON, Canada

[5]Temerty Centre for Therapeutic Brain Intervention, Centre for Addiction and Mental Health, Toronto, ON, Canada

[6]THETA Collaborative, University Health Network, University of Toronto, Toronto, ON, Canada

[7]Biostatistics Unit, University Health Network, University of Toronto, Toronto, ON, Canada

[8]Aboriginal Engagement and Outreach, Centre for Addiction and Mental Health, Toronto, ON, Canada

[9]Department of Psychology, University of Toronto, Toronto, ON, Canada

[10]Department of Psychiatry, University of California San Diego, La Jolla, CA, United States

[11]Mood and Anxiety Ambulatory Services, Centre for Addiction and Mental Health, Toronto, ON, Canada

Corresponding Author:
Zafiris Daskalakis, MD, PhD
Department of Psychiatry
University of California San Diego
9500 Gilman Drive
La Jolla, CA, 92093
United States
Phone: 1 (858) 534 2230
Email: zdaskalakis@health.ucsd.edu

Abstract

Background: Approximately 70% of mental health disorders appear prior to 25 years of age and can become chronic when ineffectively treated. Individuals between 18 and 25 years old are significantly more likely to experience mental health disorders, substance dependencies, and suicidality. Treatment progress, capitalizing on the tendencies of youth to communicate online, can strategically address depressive disorders.

Objective: We performed a randomized controlled trial (RCT) that compared online mindfulness-based cognitive behavioral therapy (CBT-M) combined with standard psychiatric care to standard psychiatric care alone in youth (18-30 years old) diagnosed with major depressive disorder.

Methods: Forty-five participants were randomly assigned to CBT-M and standard care (n=22) or to standard psychiatric care alone (n=23). All participants were provided standard psychiatric care (ie, 1 session per month), while participants in the experimental group received an additional intervention consisting of the CBT-M online software program. Interaction with online workbooks was combined with navigation coaching delivered by phone and secure text messaging.

Results: In a two-level linear mixed-effects model intention-to-treat analysis, significant between-group differences were found for the Beck Depression Inventory-II score (difference –8.54, P=.01), Quick Inventory of Depressive Symptoms score (difference –4.94, P=.001), Beck Anxiety Inventory score (difference –11.29, P<.001), and Brief Pain Inventory score (difference –1.99, P=.03), while marginal differences were found for the Five Facet Mindfulness Questionnaire–Nonjudging subscale (difference –2.68, P=.05).

Conclusions: These results confirm that youth depression can be effectively treated with online CBT-M that can be delivered with less geographic restriction.

Trial Registration: Clinical Trials.gov NCT03406052; https://www.clinicaltrials.gov/ct2/show/NCT03406052

(J Med Internet Res 2021;23(3):e24380) doi:10.2196/24380

KEYWORDS

intervention study; telemedicine; electronic CBT; clinical trial; depression; cognitive behavioral therapy; CBT; online therapy; online intervention; youth; young adult

Introduction

Approximately 70% of all mental health problems appear before 25 years of age and often become chronic when not treated or ineffectively treated [1]. Such data raise questions about elevated depression rates in youth [2] as exemplified in the National Survey on Drug Use and Health (N=611,880), which found an increase in the depressive episode rates by 63% from 2009 to 2017 [2]. Youthful online engagement makes online intervention delivery attractive, particularly with possible reductions in costs, geographic barriers, and access inequities [3-6].

Cognitive behavioral therapy (CBT) is the best-validated psychotherapy [7], and in recent years has been coordinated with mindfulness meditation (CBT-M) following strong evidence of the combined efficacy [8,9]. Online CBT-M research with student and adult populations has yielded psychometric and neurophysiological [10-20] benefits in single-arm and randomized controlled trials (RCTs). These results join a growing world literature supporting online CBT efficacy, exemplified in a meta-analysis of 3876 RCT participants indicating that online CBT was significantly more effective than control conditions in reducing depressive symptoms (Hedges g=0.27) [21]. Individual RCTs have shown online CBT to be equally effective to in-person CBT in studies with large effect sizes, along with substantial remission rates for major depressive disorder [21].

These findings motivated a focus on assessing online CBT in patients concurrently receiving standard psychiatric care to examine whether online CBT *and* psychiatric treatment as usual (TAU) was superior to psychiatric TAU. Control participants received pharmacotherapy only when deemed appropriate by treating psychiatrists, and the TAU comparison accounted for the standard use of and response to antidepressant medications.

The key behavioral intervention in this study was online access to 24 CBT workbooks and 56 mindfulness instruction videos that supported metacognitive change and autonomic balance [22], which have been linked to improved mood and reduced anxiety [23]. Navigation coaching was supplied by students who were pursuing graduate degrees (MSc, MA, PhD) in kinesiology and health science, education, and psychology. Their group training (prior to and during the study) took place at a seminar (for 1.5 hours weekly) that focused on reviews of CBT and mindfulness-based clinical research supplemented by anonymized case discussions. One (cumulative) hour of coaching was provided to each participant weekly during 24 weeks (which included text-message exchanges with participants), and each coach received 1-hour weekly sessions of one-to-one supervision.

Navigation coaching has been increasingly applied to support the adoption of evidence-based, health-related behaviors as demonstrated in adherence to cancer screening, and exercise and diet regimens [24]. Our experiences with navigation coaching include assisting patients with type 2 diabetes to reduce hemoglobin A1c blood levels [18-20] and assisting individuals in undertaking colorectal cancer screening [25]. Positive outcomes suggest that navigation coaching can be applied to the treatment of depression in assisting the use of CBT and mindfulness methods to address depressive symptoms [26]. Past successes with text messaging–assisted interventions (with critical medical outcomes) [18-20] further influenced the emphasis on text messaging between navigator coaches and patients.

Our study objective was to assess whether online CBT-M with weekly interactions with a coach navigator and standard psychiatric care was superior to standard psychiatric care alone (as workbooks and videos were never provided without coach navigator assistance). We hypothesized that intervention participants would demonstrate significant improvements in primary outcomes when compared to the waitlist controls who received only standard psychiatric TAU.

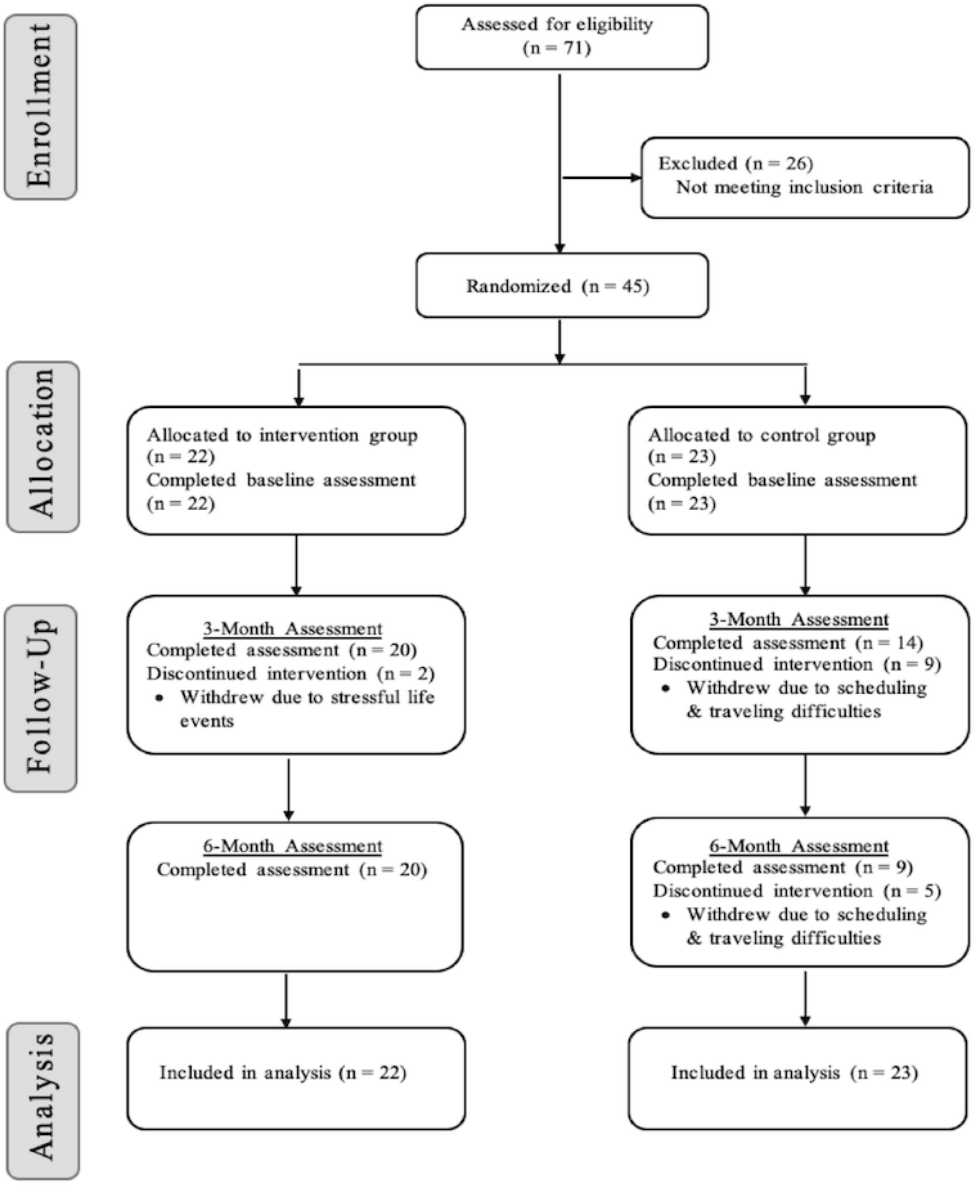
Methods

Design and Recruitment

The study was approved by the Research and Ethics Boards of the Centre for Addiction and Mental Health (CAMH; Protocol Reference number 115/2016-01) and York University (Certificate number 2017–154) in Toronto, Canada, and was registered at ClinicalTrial.gov (NCT03406052). This included distinct software platform approval for all privacy and security requirements at CAMH. The study evaluated the efficacy of CBT-M to treat young adults (18-30 years of age) with major depressive disorder. Participants were identified from service wait lists at the CAMH by research coordinators and in the prescreening of new clinic referrals. The investigative team was informed about possible participant eligibility and the client's clinician was notified. The clinician then asked the client if she/he was willing to meet with a study team member to explore participation. Information about the study was only shared once the clients agreed to meet for potential participation. A biostatistician (GT) performed electronic randomization of participants, assigning study IDs to intervention vs waitlist control participants. Information regarding each study ID with its respective group allocation was transferred onto cards placed in individually sealed, opaque envelopes. After a participant completed baseline questionnaires, the research coordinator opened the next envelope in the sequence to assign the group and respective study ID.

The inclusion criteria were: (1) aged 18 to 30 years; (2) Beck Depression Inventory-2 (BDI-II) score of at least mild severity, with no upper limit (BDI-II score≥14) [27]; (3) Mini-International Neuropsychiatric Interview (MINI)-confirmed psychiatric diagnosis of major depressive disorder [28]; and (4) fluent in English. All patients were diagnosed by a CAMH physician, with diagnoses confirmed using the MINI interview administered at the screening visit. The exclusion criteria were: (1) individuals who were currently receiving weekly structured psychotherapy; (2) individuals who met the Diagnostic and Statistical Manual of Mental Disorders-V criteria for severe alcohol/substance use disorders in the past 3 months, individuals who demonstrated clinically significant suicidal ideation (defined as imminent intent), and individuals who had attempted suicide in the past 6 months; and (3) individuals with comorbid diagnoses of borderline personality, bipolar disorder, schizophrenia, and/or obsessive compulsive disorder (Figure 1).

Figure 1. Consort flow diagram.



Intervention

All participants were provided standard psychiatric care, operationally defined as one monthly session of TAU and mostly pharmacotherapy focused. Experimental participants received the additional CBT-M program content (workbooks and videos) accessed online through a software platform developed by NexJ Health, Inc. The platform, NexJ Connected Wellness (NCW), has unique properties that facilitated participant use. Interactions with the online workbooks were combined with navigation coaching (total 24-hour duration over 6 months), delivered as phone and text message exchanges. Each participant was also given a Fitbit-HR Charge 2 that assessed physical steps and 24-hour heart rate in 5-second durations combined with the NexJ Health, Inc software that permitted daily monitoring.

The intervention content was built on two prior successful web-based CBT-M RCTs with students [10-17] and on effective methods used with other populations assessed in RCTs [29-39]. The online content included 24 workbook chapters reflecting multiple topics (eg, Living by Your Truths, Overcoming Wired-ness and Tired-ness, Mindfulness and Relationships, Loss and Grief, Resilience, Befriending Ourselves, Befriending Your Body With Exercise, Body Image and Mindfulness, Intimacy, Forgiveness, Overcoming Procrastination, Dealing With Negative Moods, Stress Resilience, Overcoming Performance Anxiety, and Cultivating Inspiration). The content was covered sequentially on a weekly basis with navigator coach guidance. In summary, key intervention features were 24-hour access and CBT-M contents that addressed specific symptoms and generic depressive experiences. The online platform used is produced and maintained by NexJ Health, Inc in Toronto, Ontario, and is the same basic platform employed in a prior study [19], although it has been upgraded numerous times in the interim. NexJ Health, Inc provided use of the NCW platform free of charge (as a research partner) but contributed no other funding or support for the study.

Outcome Measures

The primary outcome measure was the BDI-II [27], and the secondary outcomes focused on anxiety (Beck Anxiety Inventory [BAI]) [40], depression (Quick Inventory of Depressive Symptomatology [QIDS]) [41], the 24-item Hamilton Depression Rating Scale (HDRS-24; assessed by a blinded interview rater) [42], mindfulness (5-Facet Mindfulness Questionnaire [FFMQ]) [43], and pain (Brief Pain Inventory [BPI]) [44].

All self-report measures and the HDRS-24 interviews were conducted at the same CAMH Ambulatory Service setting. The HDRS-24 interview rater was blinded to intervention and control conditions for the trial duration.

Statistical Analysis

We used a two-level linear mixed-effects model to compare the difference in the rate of change regarding outcome scores between the intervention and control groups, accounting for the repeated measurement nature of the data. A full information maximum-likelihood method was used to deal with missing data [45]. Age, sex, and ethnicity were further included as auxiliary variables for this approach.

Results

Analyses

Data obtained from participants during study visits were deidentified and stored as electronic case reporting forms (CRFs) on the CAMH REDCap system, with the CRF paper copies stored in a secure, locked cabinet. Participant characteristics are summarized via descriptive statistics in Table 1. Group equivalence at baseline in terms of demographic and clinical variables was confirmed.

Table 1. Baseline demographic characteristics of study participants.

Characteristics	CBT-M[a] (n=22)	WLC[b] (n=23)	P value
Age (years), mean (SD)	25 (3.319)	24 (3.233)	.41
Gender, n (%)			.30
Male	10 (46)	7 (30)	
Female	12 (54)	16 (70)	
Ethnicity, n (%)			.57
Caucasian	13 (59)	12 (52)	
Asian	6 (27)	4 (17)	
African-American	0 (0)	1 (4)	
Indigenous	0 (0)	1 (4)	
Other	3 (14)	5 (22)	
Relationship status, n (%)			.51
Married	1 (5)	0 (0)	
Single	20 (90)	21 (91)	
Other (eg, common law)	1 (5)	2 (9)	
Offspring, n	0 (0)	0 (0)	N/A[c]
Work status, n (%)			.47
Employed	12 (55)	15 (65)	
Not employed	10 (45)	8 (35)	
Depression duration, mean (SD)			
Depression since onset age (years)	17 (4.13)	17 (5.01)	.98
Duration of current/last depressive episode (months)	9 (14.60)	19 (30.86)	.19
Number of identified depressive episodes	5.5 (5.06)	6.1 (6.94)	.72
Psychiatric history, mean			
Previous medication trials and failures	1.09	1.21	.91
Level of substance dependency or abuse	0	0	N/A
Number of suicide attempts (from MINI[d]), mean	0	0	N/A
Comorbidities, mean			
Psychiatric comorbidities	3.09	3.26	.77
Physical comorbidities	1.59	1.70	.81
Outcomes, mean (SD)			
Baseline BDI-II[e]	30 (8.40)	27 (7.90)	.21
Baseline BAI[f] (mean)	29 (8.53)	22 (9.40)	.008
Baseline BPI[g] (average pain x/10)	1.9 (2.50)	1.6 (2.25)	.64
HDRS[h]	26 (6.96)	26 (6.43)	.96
QIDS[i]	16 (4.30)	15 (3.70)	.44
FFMQ[j]-Observing	15 (3.30)	13 (3.54)	.21
FFMQ-Describe	15 (5.25)	14 (3.94)	.45
FFMQ-Act Aware	12 (3.81)	14 (3.24)	.08
FFMQ-Nonjudging	11 (4.03)	13 (3.52)	.09
FFMQ-Nonreactivity	12 (3.25)	11 (3.44)	.32

[a]CBT-M: mindfulness-based cognitive behavioral therapy.
[b]WLC: waitlist control.
[c]N/A: not applicable.
[d]MINI: Mini-International Neuropsychiatric Interview.
[e]BDI-II: Beck Depression Inventory-2.
[f]BAI: Beck Anxiety Inventory.
[g]BPI: Brief Pain Inventory.
[h]HDRS: Hamilton Depression Rating Scale.
[i]QIDS: Quick Inventory of Depressive Symptomatology.
[j]FFMQ: 5-Facet Mindfulness Questionnaire.

Primary and Secondary Outcomes

A mean of 4.7 participants were enrolled per month. The intervention and TAU retention differed markedly as 91% (20/22) of intervention participants were retained compared to only 39% (9/23) of the TAU group (at the end of the trial). Of the 14 control dropouts, 9 dropped out shortly after the baseline assessment and 5 dropped out following completion of midterm, 3-month measures. Of the 2 intervention dropouts, both dropped out shortly after enrollment (prior to 3-month measures). The between-group retention differences were significant at 3 months (P=.04) and 6 months (P=.001).

In the two-level linear model intention-to-treat analysis, between-group BDI-II, QIDS, BAI, BPI, and FFMQ–Nonjudging subscale differences were statistically significant (Table 2).

In the within-group differences, participants who completed the intervention (n=20) demonstrated significant reductions in depressive and anxiety symptoms as measured by changes in BDI-II (P<.001), BAI (P<.001), QIDS (P<.001), and (blinded) HDRS (P<.001) scores from pre- to postintervention (Table 3). The effect sizes were very large for the BDI-II (Cohen *d*=1.90 and Hedges *g*=1.82) and large on the QIDS (Cohen *d*=1.43 and Hedges *g*=1.38). All effect sizes were large and two were at or above 1.6 (Cohen *d*), typically calculated as two times a large effect size.

Table 2. Between-group differences based on intention-to-treat analysis (N=45).

Outcome	Pretreatment, mean (SD)		Final assessment, mean (SD)		Difference[a]	P value
	Intervention	Control	Intervention	Control		
BDI-II[b,c]	30.14 (8.397)	27.00 (7.909)	13.6 (9.73)	19.78 (16.642)	–8.54	.01
BAI[d,e]	29.14 (8.532)	21.74 (9.401)	15.45 (9.145)	21.67 (15.248)	–11.29	.001
QIDS[e,f]	15.50 (4.307)	14.57 (3.703)	8.95 (4.925)	12.89 (6.03)	–4.94	.001
HDRS[e,g]	26.27 (6.964)	26.17 (6.436)	14.75 (8.735)	22.67 (13.134)	–5.52	.09
BPI[e,h]	1.89 (2.465)	1.553 (2.258)	1.11 (1.753)	2.17 (2.246)	–1.99	.03
FFMQ[i]–Nonjudging[e]	11.23 (4.035)	13.17 (3.525)	14.95 (5.042)	13.56 (5.199)	2.68	.049
FFMQ–Describing[e]	15.27 (5.248)	14.22 (3.942)	18.45 (4.032)	16.11 (4.314)	0.62	.61
FFMQ–Observing[e]	14.55 (3.306)	13.26 (3.454)	15.15 (3.829)	14.22 (4.738)	–1.22	.16
FFMQ–Awareness[e]	12.33 (3.816)	14.13 (3.238)	16.2 (4.396)	15.56 (4.667)	2.16	.11

[a]Difference of rate of change: a negative value indicates greater reduction in the intervention group.
[b]BDI-II: Beck Depression Inventory-2.
[c]Planned analysis of primary outcome.
[d]BAI: Beck Anxiety Inventory.
[e]Bonferroni correction was not applied for secondary outcomes.
[f]QIDS: Quick Inventory of Depressive Symptomatology.
[g]HDRS: Hamilton Depression Rating Scale.
[h]BPI: Brief Pain Inventory.
[i]FFMQ: 5-Facet Mindfulness Questionnaire.

Table 3. Within-group differences of intervention participants who completed the trial (N=20).

Scale	Change (95% CI)	P value	Cohen *d*
BDI-II[a]	–15.6 (–2.02 to –11.1)	<.001	1.9
BAI[b]	–127 (–16.9 to –8.5)	<.001	1.5
QIDS[c]	–6.2 (–8.3 to –4.0)	<.001	1.4
HDRS[d]	–10.7 (–14.7 to –6.6)	<.001	1.6
BPI[e]	–0.8 (–1.8 to 0.3)	.14	—[f]
FFMQ[g]			
Observing	0.01 (–0.7 to 0.9)	.46	—
Describing	2.5 (0.5-4.5)	.02	1.5
Awareness	2.8 (1.3-4.3)	.001	1.4
Nonjudging	2.5 (0.9-4.2)	.005	1.4

[a]BDI-II: Beck Depression Inventory-2.

[b]BAI: Beck Anxiety Inventory.

[c]QIDS: Quick Inventory of Depressive Symptomatology.

[d]HDRS: Hamilton Depression Rating Scale.

[e]BPI: Brief Pain Inventory.

[f]—:not applicable.

[g]FFMQ: 5-Facet Mindfulness Questionnaire.

Discussion

The online CBT-M intervention was beneficial, given significant between-group differences in depression (BDI-II, QIDS), anxiety (BAI), pain (BPI), and mindfulness (FFMQ–Nonjudging subscale). Other notable between-group observations involved a 9% dropout rate in the intervention group that significantly differed from the 61% dropout rate in the TAU control group (the 61% dropout rate was estimated at ~14% above the mean for CAMH TAU). This difference suggests that the intervention had positive effects on participant retention. The intervention sample included a large subgroup with severe depression (n=10 participants, defined as severe by a BD-II score>29), 50% of whom were in remission (BDI<14) at the final (6-month) assessment. Of the 6 participants who exhibited moderate depression, 5 achieved remission, and of the 3 study participants with mild depression, 2 achieved remission.

Given the CBT-M intervention, it was notable that between-group differences were found in the Nonjudging subscale of the FFMQ that assesses the excess self-critical thinking associated with distress [43]. Intervention group participants engaged in significantly less self-critical self-judgment at the 6-month follow up than the TAU controls. Although the study sample size did not allow for mediation analyses [46], the between-group difference observed suggests that the mindfulness component of the CBT-M intervention was likely involved in the modification of depressogenic cognitions [2]. The between-group differences also appear linked to the self-acceptance emphasis in the CBT-M interventions employed.

Significant between-group differences were found in self-reported chronic pain as indicated in the BPI scale. The inclusion of a pain assessment reflects recent findings about the high comorbidity prevalence in depression with respect to chronic pain [47] and the possible efficacy of behavioral pain reduction methods [48,49]. Despite this study's findings, the behavioral intervention literature on reductions in chronic pain remains sparse, and additional targeted studies in populations with pain and mental health difficulties are warranted.

Although attrition in psychiatric treatment has been linked to early improvements associated with medication changes [50], this explanation does not seem to apply in the current trial, as 4 of the 5 total control participants who discontinued after the 3-month midterm measures received no pharmacotherapy or no modifications in the prestudy pharmacotherapy established. The control participants who dropped out following the baseline assessment (before midterm, 3-month assessment; n=9) did not receive medication initiation or modification. The attrition difference also did not appear to be based on more severe baseline depression as the mean BDI-II depression score at baseline for TAU control participants (BDI-II=27.0) reflected milder depression symptoms than those of intervention participants (mean BDI-II=30.14).

Significant study strengths included the control comparison with a standard-care psychiatry intervention, delivered at the same institution, independently versus in combination with the experimental behavioral intervention. This resulted in detailed records of how pharmacological and behavioral interventions interacted, assisting estimations of independent and combined benefits. The study further controlled for the intervention-related placebo effects observed in 35%-40% of RCT participants exposed to TAU conditions [7]. This was also a necessary control for medication effects, given that individuals treated for depression show improvement with antidepressants alone [7]. In this study, CBT-M effects were clearly additive to TAU

effects. Although the TAU-only group attrition rate can be seen as a study limitation, there were demonstrated associations between the CBT-M intervention and retention (ie, lower attrition in the experimental group) that indicated retention benefits associated with the behavioral treatment.

In recent meta-analyses focused on CBT-M delivery for depression, multiple CBT modalities have been assessed, notably individual, group, telephone-based, and guided self-help, all of which appear to be significantly more effective than waitlist and care-as-usual control conditions, and unguided self-help [51]. These analyses reflect the investigative search for the most cost-effective CBT delivery. In the context of current meta-analyses, our intervention can be characterized as combining telephone-based with guided self-help (online), with results that show significantly better outcomes than care-as-usual controls.

Key limitations of our study include a lack of participant blinding and the limited power associated with a small sample size. However, the HDRS assessment was undertaken by a rater blinded to group allocation. Although the between-group differences on the HDRS were trending toward significance (P=.09), they were not statistically significant. A final limitation is that the study psychiatrists administering TAU to control and intervention participants were not blind to which participants were in the intervention versus control groups, and this might have led to biased treatment.

Future studies comparing CBT-M and standard-care psychiatry would benefit from larger sample sizes, more complete blinding, and extended follow up after intervention conclusion (eg, 6-12 months). Despite these limitations, the results indicate that online CBT-M combined with TAU psychiatric treatment was an effective treatment for major depressive disorder and led to significantly greater reductions in BDI-II scores than TAU psychiatry alone.

Acknowledgments

PR has received research support from the Canadian Institutes of Health Research (CIHR) and the Federal Development Program of Southern Ontario. ZD has received research support from CIHR, National Institutes of Health, Brain Canada, and the Temerty family through the CAMH Foundation and the Campbell Family Research Institute. JK is supported by a Canadian Institutes of Health Research Canada Research Chair in Health Psychology at York University.

Conflicts of Interest

ZD received research support and in-kind equipment support for an investigator-initiated study from Brainsway, Ltd. He is the site principal investigator for sponsor-initiated studies for Brainsway, Ltd. He also receives in-kind equipment support Magventure for investigator-initiated research. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 733 KB - jmir_v23i3e24380_app1.pdf]

References

1. Government of Canada. The Human Face of Mental Health and Mental Illness in Canada 2006. Ottawa, Ontario: Minister of Public Works and Government Services; 2006. URL: http://www.phac-aspc.gc.ca/publicat/human-humain06/pdf/human_face_e.pdf [accessed 2021-02-05]
2. Twenge JM, Cooper AB, Joiner TE, Duffy ME, Binau SG. Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005-2017. J Abnorm Psychol 2019 Apr;128(3):185-199. [doi: 10.1037/abn0000410] [Medline: 30869927]
3. McCrone P, Knapp M, Proudfoot J, Ryden C, Cavanagh K, Shapiro DA, et al. Cost-effectiveness of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. Br J Psychiatry 2004 Jul 02;185(1):55-62. [doi: 10.1192/bjp.185.1.55] [Medline: 15231556]
4. Warmerdam L, Smit F, van Straten A, Riper H, Cuijpers P. Cost-utility and cost-effectiveness of internet-based treatment for adults with depressive symptoms: randomized trial. J Med Internet Res 2010 Dec 19;12(5):e53 [FREE Full text] [doi: 10.2196/jmir.1436] [Medline: 21169166]
5. Hollinghurst S, Peters TJ, Kaur S, Wiles N, Lewis G, Kessler D. Cost-effectiveness of therapist-delivered online cognitive-behavioural therapy for depression: randomised controlled trial. Br J Psychiatry 2010 Oct 02;197(4):297-304. [doi: 10.1192/bjp.bp.109.073080] [Medline: 20884953]
6. Romero-Sanchiz P, Nogueira-Arjona R, García-Ruiz A, Luciano JV, García Campayo J, Gili M, et al. Economic evaluation of a guided and unguided internet-based CBT intervention for major depression: Results from a multi-center, three-armed randomized controlled trial conducted in primary care. PLoS One 2017 Feb 27;12(2):e0172741 [FREE Full text] [doi: 10.1371/journal.pone.0172741] [Medline: 28241025]

7. Hofmann SG, Smits JAJ. Cognitive-behavioral therapy for adult anxiety disorders: a meta-analysis of randomized placebo-controlled trials. J Clin Psychiatry 2008 Apr 15;69(4):621-632 [FREE Full text] [doi: 10.4088/jcp.v69n0415] [Medline: 18363421]
8. Sipe WEB, Eisendrath SJ. Mindfulness-based cognitive therapy: theory and practice. Can J Psychiatry 2012 Feb;57(2):63-69. [doi: 10.1177/070674371205700202] [Medline: 22340145]
9. Coelho HF, Canter PH, Ernst E. Mindfulness-based cognitive therapy: evaluating current evidence and informing future research. J Consult Clin Psychol 2007 Dec;75(6):1000-1005. [doi: 10.1037/0022-006X.75.6.1000] [Medline: 18085916]
10. Ritvo P, Vora K, Irvine J, Mongrain M, Azam A, Azargive S. Reductions in negative automatic thoughts in students attending mindfulness tutorial predicts increased life satisfaction. Int J Educ Psychol 2013 Oct;2(3):272-296. [doi: 10.4471/ijep.2013.28]
11. Azam MA, Mongrain M, Vora K, Pirbaglou M, Azargive S, Changoor T, et al. Mindfulness as an alternative for supporting university student mental health: cognitive-emotional and depressive self-criticism measures. Int J Educ Psychol 2016 Jun 24;5(2):140. [doi: 10.17583/ijep.2016.1504]
12. Guglietti CL, Daskalakis ZJ, Radhu N, Fitzgerald PB, Ritvo P. Meditation-related increases in GABAB modulated cortical inhibition. Brain Stimul 2013 May;6(3):397-402. [doi: 10.1016/j.brs.2012.08.005] [Medline: 23022436]
13. Radhu N, Daskalakis ZJ, Guglietti CL, Farzan F, Barr MS, Arpin-Cribbie CA, et al. Cognitive behavioral therapy-related increases in cortical inhibition in problematic perfectionists. Brain Stimul 2012 Jan;5(1):44-54. [doi: 10.1016/j.brs.2011.01.006] [Medline: 22037137]
14. Radhu N, Daskalakis ZJ, Arpin-Cribbie CA, Irvine J, Ritvo P. Evaluating a web-based cognitive-behavioral therapy for maladaptive perfectionism in university students. J Am Coll Health 2012 Jul;60(5):357-366. [doi: 10.1080/07448481.2011.630703] [Medline: 22686358]
15. Arpin-Cribbie C, Irvine J, Ritvo P. Web-based cognitive-behavioral therapy for perfectionism: a randomized controlled trial. Psychother Res 2012 Mar;22(2):194-207. [doi: 10.1080/10503307.2011.637242] [Medline: 22122217]
16. Brief Summary of the Inaugural 12 Day Mindfulness Challenge. The Canadian Clinical Psychologist spring newsletter. 2015 Apr. URL: https://cpa.ca/docs/File/Sections/Clinical%20section/newsletters/Spring2015newsletter.pdf [accessed 2021-02-05]
17. Azam MA, Katz J, Fashler SR, Changoor T, Azargive S, Ritvo P. Heart rate variability is enhanced in controls but not maladaptive perfectionists during brief mindfulness meditation following stress-induction: A stratified-randomized trial. Int J Psychophysiol 2015 Oct;98(1):27-34. [doi: 10.1016/j.ijpsycho.2015.06.005] [Medline: 26116778]
18. Wayne N, Ritvo P. Smartphone-enabled health coach intervention for people with diabetes from a modest socioeconomic strata community: single-arm longitudinal feasibility study. J Med Internet Res 2014 Jun 06;16(6):e149 [FREE Full text] [doi: 10.2196/jmir.3180] [Medline: 24907918]
19. Wayne N, Perez DF, Kaplan DM, Ritvo P. Health coaching reduces HbA1c in type 2 diabetic patients from a lower-socioeconomic status community: a randomized controlled trial. J Med Internet Res 2015 Oct 05;17(10):e224 [FREE Full text] [doi: 10.2196/jmir.4871] [Medline: 26441467]
20. Pludwinski S, Ahmad F, Wayne N, Ritvo P. Participant experiences in a smartphone-based health coaching intervention for type 2 diabetes: A qualitative inquiry. J Telemed Telecare 2016 Apr 21;22(3):172-178. [doi: 10.1177/1357633X15595178] [Medline: 26199275]
21. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
22. Conklin QA, Crosswell AD, Saron CD, Epel ES. Meditation, stress processes, and telomere biology. Curr Opin Psychol 2019 Aug;28:92-101 [FREE Full text] [doi: 10.1016/j.copsyc.2018.11.009] [Medline: 30553080]
23. Ramel W, Goldin PR, Carmona PE, McQuaid JR. The effects of mindfulness meditation on cognitive processes and affect in patients with past depression. Cogn Ther Res 2004 Aug;28(4):433-455. [doi: 10.1023/B:COTR.0000045557.15923.96]
24. Cramer H, Lauche R, Paul A, Dobos G. Mindfulness-based stress reduction for breast cancer-a systematic review and meta-analysis. Curr Oncol 2012 Oct;19(5):e343-e352 [FREE Full text] [doi: 10.3747/co.19.1016] [Medline: 23144582]
25. Ritvo PG, Myers RE, Paszat LF, Tinmouth JM, McColeman J, Mitchell B, et al. Personal navigation increases colorectal cancer screening uptake. Cancer Epidemiol Biomarkers Prev 2015 Mar;24(3):506-511 [FREE Full text] [doi: 10.1158/1055-9965.EPI-14-0744] [Medline: 25378365]
26. Almeida OP, Marsh K, Murray K, Hickey M, Sim M, Ford A, et al. Reducing depression during the menopausal transition with health coaching: Results from the healthy menopausal transition randomised controlled trial. Maturitas 2016 Oct;92:41-48. [doi: 10.1016/j.maturitas.2016.07.012] [Medline: 27621237]
27. Smarr KL, Keefer AL. Measures of depression and depressive symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9). Arthritis Care Res (Hoboken) 2011 Nov 07;63(Suppl 11):S454-S466. [doi: 10.1002/acr.20556] [Medline: 22588766]

28. Sheehan D, Lecrubier Y, Sheehan K, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59(Suppl 20):22-33;quiz 34. [Medline: 9881538]
29. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
30. Boettcher J, Aström V, Påhlsson D, Schenström O, Andersson G, Carlbring P. Internet-based mindfulness treatment for anxiety disorders: a randomized controlled trial. Behav Ther 2014 Mar;45(2):241-253 [FREE Full text] [doi: 10.1016/j.beth.2013.11.003] [Medline: 24491199]
31. Carlbring P, Hägglund M, Luthström A, Dahlin M, Kadowaki N, Vernmark K, et al. Internet-based behavioral activation and acceptance-based treatment for depression: a randomized controlled trial. J Affect Disord 2013 Jun;148(2-3):331-337. [doi: 10.1016/j.jad.2012.12.020] [Medline: 23357657]
32. Dahlin M, Andersson G, Magnusson K, Johansson T, Sjögren J, Håkansson A, et al. Internet-delivered acceptance-based behaviour therapy for generalized anxiety disorder: A randomized controlled trial. Behav Res Ther 2016 Feb;77:86-95. [doi: 10.1016/j.brat.2015.12.007] [Medline: 26731173]
33. Gershkovich M, Herbert JD, Forman EM, Schumacher LM, Fischer LE. Internet-delivered acceptance-based cognitive-behavioral intervention for social anxiety disorder with and without therapist support: a randomized trial. Behav Modif 2017 Sep 08;41(5):583-608. [doi: 10.1177/0145445517694457] [Medline: 28776431]
34. Kivi M, Eriksson MCM, Hange D, Petersson E, Vernmark K, Johansson B, et al. Internet-based therapy for mild to moderate depression in Swedish primary care: short term results from the PRIM-NET randomized controlled trial. Cogn Behav Ther 2014;43(4):289-298 [FREE Full text] [doi: 10.1080/16506073.2014.921834] [Medline: 24911260]
35. Lappalainen P, Granlund A, Siltanen S, Ahonen S, Vitikainen M, Tolvanen A, et al. ACT Internet-based vs face-to-face? A randomized controlled trial of two ways to deliver Acceptance and Commitment Therapy for depressive symptoms: an 18-month follow-up. Behav Res Ther 2014 Oct;61:43-54. [doi: 10.1016/j.brat.2014.07.006] [Medline: 25127179]
36. Lappalainen P, Langrial S, Oinas-Kukkonen H, Tolvanen A, Lappalainen R. Web-based acceptance and commitment therapy for depressive symptoms with minimal support: a randomized controlled trial. Behav Modif 2015 Nov;39(6):805-834. [doi: 10.1177/0145445515598142] [Medline: 26253644]
37. Ly KH, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, et al. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open 2014 Jan 09;4(1):e003440 [FREE Full text] [doi: 10.1136/bmjopen-2013-003440] [Medline: 24413342]
38. Pots WTM, Fledderus M, Meulenbeek PAM, ten Klooster PM, Schreurs KMG, Bohlmeijer ET. Acceptance and commitment therapy as a web-based intervention for depressive symptoms: randomised controlled trial. Br J Psychiatry 2016 Jan;208(1):69-77. [doi: 10.1192/bjp.bp.114.146068] [Medline: 26250745]
39. Silfvernagel K, Carlbring P, Kabo J, Edström S, Eriksson J, Månson L, et al. Individually tailored internet-based treatment for young adults and adults with panic attacks: randomized controlled trial. J Med Internet Res 2012 Jun 26;14(3):e65 [FREE Full text] [doi: 10.2196/jmir.1853] [Medline: 22732098]
40. Julian LJ. Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). Arthritis Care Res (Hoboken) 2011 Nov;63(Suppl 11):S467-S472. [doi: 10.1002/acr.20561] [Medline: 22588767]
41. Reilly TJ, MacGillivray SA, Reid IC, Cameron IM. Psychometric properties of the 16-item Quick Inventory of Depressive Symptomatology: a systematic review and meta-analysis. J Psychiatr Res 2015 Jan;60:132-140. [doi: 10.1016/j.jpsychires.2014.09.008] [Medline: 25300442]
42. Worboys M. The Hamilton Rating Scale for Depression: The making of a "gold standard" and the unmaking of a chronic illness, 1960-1980. Chronic Illn 2013 Sep 21;9(3):202-219 [FREE Full text] [doi: 10.1177/1742395312467658] [Medline: 23172888]
43. Williams MJ, Dalgleish T, Karl A, Kuyken W. Examining the factor structures of the five facet mindfulness questionnaire and the self-compassion scale. Psychol Assess 2014 Jun;26(2):407-418. [doi: 10.1037/a0035566] [Medline: 24490681]
44. Lapane KL, Quilliam BJ, Benson C, Chow W, Kim M. One, two, or three? Constructs of the brief pain inventory among patients with non-cancer pain in the outpatient setting. J Pain Symptom Manage 2014 Feb;47(2):325-333 [FREE Full text] [doi: 10.1016/j.jpainsymman.2013.03.023] [Medline: 23880588]
45. Dong Y, Peng CYJ. Principled missing data methods for researchers. Springerplus 2013 Dec;2(1):222 [FREE Full text] [doi: 10.1186/2193-1801-2-222] [Medline: 23853744]
46. Pan H, Liu S, Miao D, Yuan Y. Sample size determination for mediation analysis of longitudinal data. BMC Med Res Methodol 2018 Mar 27;18(1):32 [FREE Full text] [doi: 10.1186/s12874-018-0473-2] [Medline: 29580203]
47. Dhanju S, Kennedy S, Abbey S, Katz J, Weinrib A, Clarke H, et al. The impact of comorbid pain and depression in the United States: results from a nationally representative survey. Scand J Pain 2019 Apr 24;19(2):319-325. [doi: 10.1515/sjpain-2018-0323] [Medline: 30759071]
48. Salomons TV, Moayedi M, Erpelding N, Davis KD. A brief cognitive-behavioural intervention for pain reduces secondary hyperalgesia. Pain 2014 Aug;155(8):1446-1452. [doi: 10.1016/j.pain.2014.02.012] [Medline: 24569149]

49. Ehde DM, Dillworth TM, Turner JA. Cognitive-behavioral therapy for individuals with chronic pain: efficacy, innovations, and directions for research. Am Psychol 2014;69(2):153-166. [doi: 10.1037/a0035747] [Medline: 24547801]
50. Rush AJ, Beck AT, Kovacs M, Hollon S. Comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. Cogn Ther Res 1977 Mar;1(1):17-37. [doi: 10.1007/bf01173502]
51. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. JAMA Psychiatry 2019 Jul 01;76(7):700-707 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.0268] [Medline: 30994877]

Abbreviations

BAI: Beck Anxiety Inventory
BDI: Beck Depression Inventory
BPI: Brief Pain Inventory
CAMH: Centre for Addiction and Mental Health
CBT: cognitive behavioral therapy
CBT-M: mindfulness-based cognitive behavioral therapy
CIHR: Canadian Institutes of Health Research
CRF: case reporting form
FFMQ: 5-Facet Mindfulness Questionnaire
HDRS-24: 24-item Hamilton Depression Rating Scale
MINI: Mini-International Neuropsychiatric Interview
NCW: NexJ Connected Wellness
QIDS: Quick Inventory of Depressive Symptomatology
RCT: randomized controlled trial
TAU: treatment as usual

Edited by R Kukafka; submitted 21.09.20; peer-reviewed by S Doan, N Wayne; comments to author 09.10.20; revised version received 30.11.20; accepted 10.01.21; published 10.03.21.

Please cite as:
Ritvo P, Knyahnytska Y, Pirbaglou M, Wang W, Tomlinson G, Zhao H, Linklater R, Bai S, Kirk M, Katz J, Harber L, Daskalakis Z
Online Mindfulness-Based Cognitive Behavioral Therapy Intervention for Youth With Major Depressive Disorders: Randomized Controlled Trial
J Med Internet Res 2021;23(3):e24380
URL: https://www.jmir.org/2021/3/e24380
doi:10.2196/24380
PMID:33688840

Original Paper

Electronic Health Risk Behavior Screening With Integrated Feedback Among Adolescents in Primary Care: Randomized Controlled Trial

Laura Richardson[1,2], MPH, MD; Elizabeth Oshrin Parker[1], PhD; Chuan Zhou[1,2], PhD; Julie Kientz[3], PhD; Elizabeth Ozer[4,5], PhD; Carolyn McCarty[1,2], PhD

[1]Seattle Children's Research Institute, Seattle, WA, United States

[2]Department of Pediatrics, University of Washington, Seattle, WA, United States

[3]Department of Human Centered Design & Engineering, University of Washington, Seattle, WA, United States

[4]Division of Adolescent and Young Adult Medicine, Department of Pediatrics, University of California, San Francisco, San Francisco, CA, United States

[5]Office of Diversity & Outreach, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Laura Richardson, MPH, MD
Seattle Children's Research Institute
1900 9th Ave
Seattle, WA, 98101
United States
Phone: 1 (206) 884 7300
Email: laura.richardson@seattlechildrens.org

Abstract

Background: Health risk behaviors are the most common sources of morbidity among adolescents. Adolescent health guidelines (Guidelines for Preventive Services by the AMA and Bright Futures by the Maternal Child Health Bureau) recommend screening and counseling, but the implementation is inconsistent.

Objective: This study aims to test the efficacy of electronic risk behavior screening with integrated patient-facing feedback on the delivery of adolescent-reported clinician counseling and risk behaviors over time.

Methods: This was a randomized controlled trial comparing an electronic tool to usual care in five pediatric clinics in the Pacific Northwest. A total of 300 participants aged 13-18 years who attended a well-care visit between September 30, 2016, and January 12, 2018, were included. Adolescents were randomized after consent by employing a 1:1 balanced age, sex, and clinic stratified schema with 150 adolescents in the intervention group and 150 in the control group. Intervention adolescents received electronic screening with integrated feedback, and the clinicians received a summary report of the results. Control adolescents received usual care. Outcomes, assessed via online survey methods, included adolescent-reported receipt of counseling during the visit (measured a day after the visit) and health risk behavior change (measured at 3 and 6 months after the visit).

Results: Of the original 300 participants, 94% (n=282), 94.3% (n=283), and 94.6% (n=284) completed follow-up surveys at 1 day, 3 months, and 6 months, respectively, with similar levels of attrition across study arms. The mean risk behavior score at baseline was 2.86 (SD 2.33) for intervention adolescents and 3.10 (SD 2.52) for control adolescents (score potential range 0-21). After adjusting for age, gender, and random effect of the clinic, intervention adolescents were 36% more likely to report having received counseling for endorsed risk behaviors than control adolescents (adjusted rate ratio 1.36, 95% CI 1.04 to 1.78) 1 day after the well-care visit. Both the intervention and control groups reported decreased risk behaviors at the 3- and 6-month follow-up assessments, with no significant group differences in risk behavior scores at either time point (3-month group difference: β=−.15, 95% CI −0.57 to −0.01, *P*=.05; 6-month group difference: β=−.12, 95% CI −0.29 to 0.52, *P*=.57).

Conclusions: Although electronic health screening with integrated feedback improves the delivery of counseling by clinicians, the impact on risk behaviors is modest and, in this study, not significantly different from usual care. More research is needed to identify effective strategies to reduce risk in the context of well-care.

Trial Registration: ClinicalTrials.gov NCT02882919; https://clinicaltrials.gov/ct2/show/NCT02882919

(J Med Internet Res 2021;23(3):e24135) doi:10.2196/24135

KEYWORDS

adolescent health services; primary care

Introduction

Background

Health risk behaviors, such as alcohol use, risky sexual behaviors, and low physical activity, are among the most common causes of morbidity and mortality during adolescence and young adulthood [1,2]. To reduce risk and morbidity, adolescent preventive care guidelines recommend screening and counseling to reduce these behaviors as a component of annual well-care visits [3,4]. However, the delivery of preventive screening is inconsistent, and only a small proportion of screened adolescents report having received counseling to reduce risk with rates of counseling varying by type of behavior [2,5-7].

Research has shown that the use of standardized screening methods, including electronic screening tools, can increase screening delivery, detection of risk, and adolescent-reported clinician counseling [6,8,9]. Adolescents report greater comfort in disclosing behaviors with electronic screening methods compared with other methods [10-13]. However, few studies have examined the impact of increasing clinician counseling on adolescent behavior outcomes in the context of multi-risk screening, as is commonly performed in well-care visits. In a recent review article examining multi-risk screening in adolescents, 9 studies were identified, with some demonstrating effects on risk behaviors [7]. Among these trials, variations in intervention duration, intensity, behaviors studied, and impacted outcomes led to a limited ability to draw definitive conclusions. In addition, based on the studies in this review, the magnitude of behavioral changes was small to modest, and the only risk behavior for which change was found in more than one study was for an increase in bicycle helmet use.

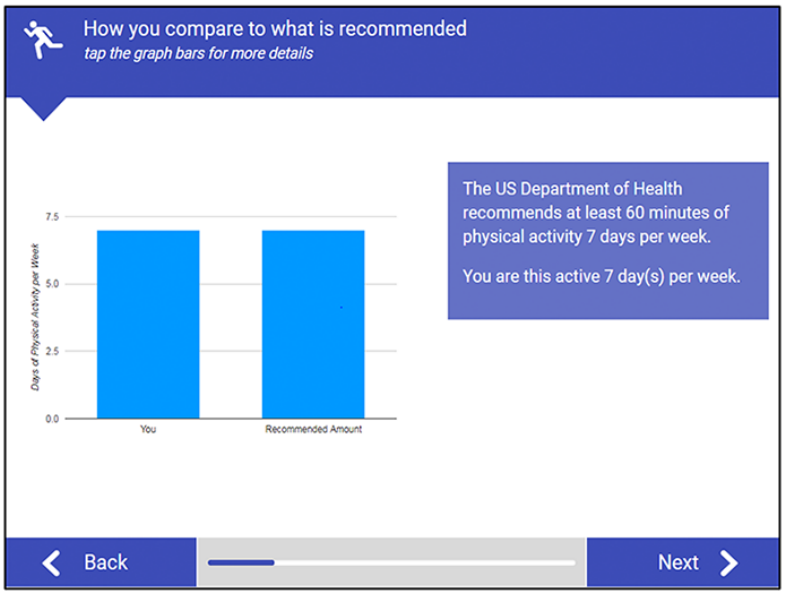
When administered consistently, electronic screening can serve to reduce biases related to selecting who gets screened and how questions are asked [14]. In addition, some studies suggest that adolescents are more likely to use preventive health services when they are given information electronically about health behaviors [15-18]. Furthermore, 2 studies found that the use of electronic screening improved adolescents' perceptions of clinician communication and partnership [19,20].

Study Aim

In this study, we aim to examine the efficacy of a tool delivered via an app or website link that combined electronic screening with integrated personalized motivational feedback. The interactive tool was developed with adolescents' input and designed to be administered before a well-care visit to prepare adolescents to discuss risk behavior change with their clinicians when indicated. The tool tested in this study (Figure 1; see additional examples in Multimedia Appendix 1) is a modified version of a previously tested tool [21]. The modifications were made to increase youth engagement with the tool and increase the ease with which clinicians can interpret the results with changes based on adolescent and clinician input gathered through a human-centered design process [22]. The tool also generated a printed one-page clinician summary of adolescent-reported behaviors. The primary outcomes of interest were adolescent-reported clinician counseling during the visit, health risk behaviors at 3 months, and patient satisfaction. The secondary outcome was health risk behavior at 6 months. We hypothesized that the intervention would increase clinician counseling and reduce health risk behaviors at 3 months.

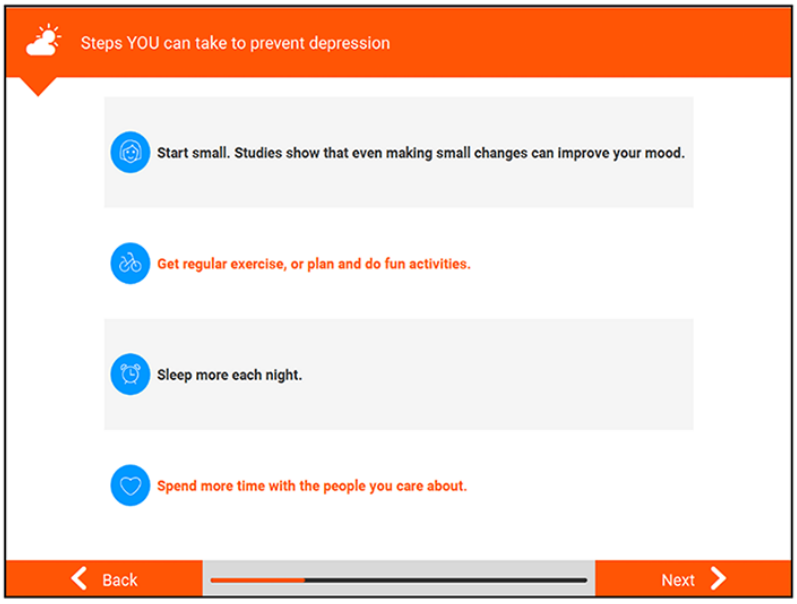
Figure 1. Screenshots of the personalized feedback from the Check Yourself tool by behavior.

Physical Activity



Copyright: Tickit Health 2016

Depression



Copyright: Tickit Health 2016

Methods

Overview

We conducted a parallel-group randomized controlled intervention study comparing the electronic screening and feedback tool to usual care in the context of a well-care visit. Adolescent participants (aged 13-18 years) were recruited from 5 pediatric clinics in Washington State between September 30, 2016, and January 12, 2018. Clinics were contacted via the Puget Sound Pediatric Research Network and invited to participate based on interest in the study and the number of adolescent patients served. Clinics were located in urban, suburban, and small city locations. The providers of these clinics included physicians and advanced practitioners. Residents and other trainees were not included in the study because of concerns about continuity in the clinical setting. Among participating clinics, the average monthly number of adolescents aged 13 to 18 years with scheduled well-visits was 49 (SD 17; range 31-71).

Sites were added to the study on a rolling basis with the goal of recruiting a minimum of 60 adolescents per site for a total recruitment goal of 300 adolescents. The study sample size was predetermined by the study statistician with the goal of having 80% power to detect an effect size as small as 0.3. Enrollment goals were met and exceeded at 4 of the 5 participating clinics, resulting in a total sample of 301 adolescents. The fifth clinic began enrollment late and, after entering the study, determined that they were not comfortable sharing patient contact data for research team outreach in the manner approved by the

institutional review board. As a result, only a small number of patients were invited to participate, of whom 2 were enrolled.

Study outreach procedures included clinics sharing contact information for all adolescent patients between the ages of 13 and 17 years who were scheduled for an upcoming well-care visit. The study staff coordinated the mailing of a letter to parents from the clinic, inviting eligible adolescents to participate in the study and providing a phone number to opt out of further contact. This letter was followed by a phone call from the study staff to provide further information and assess eligibility. Exclusion criteria included planning to cancel the well-care visit, being out of the study age range, having a sibling who was previously invited to participate, lacking phone or internet access, or if the adolescent did not speak English.

Parental consent and adolescent assent were obtained via phone for participants aged 13-17 years, whereas participants aged 18 years provided direct consent. The consent forms stated that the study would compare electronic screening with feedback and provision of the results to the clinician to electronic screening alone (with no feedback or results provided to the clinician). Although most of the adolescents approached for the study spoke English, some parents did not. To support the inclusion of these adolescents, recruitment and parental consent materials were translated into Spanish and Somali, the most common languages spoken in participating clinics.

Before beginning recruitment at each clinic, the study statistician developed a computer-generated list of random numbers that was entered into REDCap [23] with a 1:1 allocation schema stratified by age (13-15 or 16-18 years), male or female gender (as provided by the study clinics based on their records), and clinic. Participant randomization occurred after the completion of consent and assent procedures and before data collection. Adolescents were not told their study assignment but could potentially determine it based on whether or not they received feedback as part of the baseline assessment. After consent was completed, adolescent participants were sent an online link for their baseline screening assessment, with or without integrated feedback, based on the study assignment. Baseline data were collected online with phone support by trained study staff before the well-care visit. As part of the procedures, control and intervention adolescents were instructed to complete their respective baseline screening components in a private setting where they could respond confidentially. All procedures for recruitment were approved by the Seattle Children's Institutional Review Board before starting study activities. The study protocol is available upon request from the corresponding author.

Intervention Procedures

Intervention adolescents completed electronic screening with integrated personalized feedback, and their clinician received a printed one-page summary report of the screening results. The electronic screening tool assessed protective factors and risk behaviors using a HEADSS pneumonic (Home, Education, Activities, Drugs, Depression, Sexuality, and Safety) framework [24]. The tool was also screened for specific nutritional behaviors (sweetened beverage intake and fruit and vegetable intake), physical activity, and sleep. The integrated feedback component was designed to deliver messages that increased motivation and self-efficacy for healthy behavior. Feedback content varied according to behavior assessed and the youth-reported risk level. It included a combination of education, tips for change, and motivational messaging, including positive reinforcement for adolescents who did not engage in risks and messages to motivate behavior change when risks were present using a combination of normative feedback comparing adolescent-reported risks to peer reports, guidelines, and goal setting.

The tool in this study is an adapted version of the Check Yourself tool (version 2) [21,25], revised to increase interactive features with input from adolescent users, clinicians in collaboration with faculty, and researchers in human-centered design. Specific changes that were developed with adolescent input include increasing image-based feedback versus text, adding functionality to allow participants to choose to see more versus less information on each topic, and the option to receive more information about topics of interest in the form of a one-time text or email. In addition, screening content was modified to add response options related to gender identity, remove screen time assessments, enhance screening for depression and anxiety, and enhance screening and new feedback related to marijuana use. On the basis of internal tracking data, the tool took an average of 15 (SD 8) minutes to complete among control adolescents and 18 (SD 10) minutes among intervention adolescents who also received integrated feedback.

The one-page paper clinician summary included a dashboard with flags categorizing the adolescent health risks as low, moderate, or high within 6 areas: nutrition, activity, substance use, emotions, sexual activity, and safety. Individual screening responses were provided below the dashboard for each area so that clinicians could examine which specific behaviors resulted in a flag. Risk behavior severity categories (high, moderate, and low) were defined a priori based on health guidelines or expert consensus (Multimedia Appendix 2) and integrated into the electronic screening algorithms. The study staff coordinated with each clinic to develop protocols so that clinicians would receive the summary report before the visit.

Control Procedures

Control adolescents completed the electronic screening portion of the tool as a baseline assessment but did not receive integrated feedback. The clinicians did not receive any screening results.

From the outset of the study, clinicians were instructed to continue their standard health risk screening procedures for all patients (intervention and control). The standard processes for all 5 sites included a combination of paper intake screeners and interviews during the visit to assess risk, but the content of the paper screeners varied. One clinic used standardized paper anxiety and depression screens. Another clinic used a self-designed form that asked about sleep and safety risks, including texting while driving, driving under influence, helmet use, and seatbelt use. Outside of these 2 examples, there was no overlap between the health risk behavior content in the electronic screening tool and the standard screening forms employed by study clinics. None of the clinics employed a standard form to screen for confidential health risk behaviors,

such as sexual activity or drug use. All of the clinics indicated that their providers asked about confidential risk behaviors during the well-care visit, although data were not available regarding the consistency of these practices.

Before enrollment, all clinicians received an invitation to complete a 15-minute online training module to orient them to the electronic tool and how to interpret the clinician summary. As randomization was at the patient level, clinicians could be exposed to both intervention and control patients.

Surveys

The baseline assessment consisted of responses from the electronic screening tool (with or without feedback depending on assignment) conducted before the well-visit (details provided in Multimedia Appendix 3). In addition, all adolescents completed online follow-up surveys 1 day, 3 months, and 6 months following their well-care visit. The 1-day follow-up survey assessed the content of the visit, including the delivery of counseling to change behavior for each screened behavior. Items assessing the visit were adapted from the Adolescent Report of the Visit developed by Ozer et al [26]. The 3-month and 6-month follow-up surveys assessed the same health risk behaviors as at baseline, collected via an online survey tool, REDCap [23]. Participants were asked about suicidality at baseline and at all follow-up time points. To ensure safety, study investigators, who are also clinicians, followed up with all participants in either study arm who reported having thoughts of harming themselves in the past 2 weeks and thoughts of killing themselves or suicide attempts in the past 3 months and assisted them in accessing clinical services.

Analysis

All data analyses were conducted using *R 3.5.0* [27] using an intent-to-treat framework. We first conducted bivariate analyses to evaluate differences in demographics and baseline risk between adolescents in the control and intervention group. Subsequently, we conducted our main analyses on the 3 primary outcome measures: clinician counseling during the visit, a summary score of health risk behaviors measured at 3 months after the visit, and patient satisfaction. Our secondary outcome measure, the health risk behaviors summary score at 6 months was analyzed together with the 3-month summary score using repeated measures analysis.

On the basis of the study design, missing data only occurred during the outcome assessments. We compared the baseline characteristics of participants with and without missing outcomes and found no differences between the groups. We further conducted sensitivity analyses for each of our primary outcomes using multiple imputation with chained equations (MICE) methods using linear regression and predictive mean matching for continuous outcomes. For categorical outcomes, we applied classification and regression tree methods for imputation using MICE methods. Estimates from the fitted models on multiple imputed data sets were pooled to generate the final results for inference. In conducting these sensitivity analyses, we found that the results were almost identical for the imputed and complete case analysis. Thus, only the complete case analysis results are presented in this paper.

Clinician Counseling Outcome

Clinician counseling during the visit, measured on the 1-day assessment, was defined as adolescent report of the clinician having counseled them to change an endorsed behavior toward health. This measure was constructed by summing all endorsed moderate- and high-risk behaviors for which adolescents reported receiving counseling. We conducted an adjusted analysis using a mixed effects Poisson regression model in which the dependent variable was the counseling measure, and the treatment group was the predictor of interest. Baseline age and sex were included as covariates, and a clinic-specific random effect was included to account for clustering within the clinic. The total number of endorsed moderate- and high-risk behaviors was entered as an offset to ensure that the regression coefficients had proper rate interpretation. As an exploratory subanalysis to evaluate whether higher-risk behaviors were more likely to receive counseling than moderate-risk behaviors, we also conducted 2 additional regression analyses focused specifically on counseling for each category of risk behaviors: high-risk and moderate-risk behaviors, controlling for the same variables as the main analysis.

Risk Behavior Outcome

The risk behavior outcome analyses employed a summary score of all assessed behaviors at 3 months (primary outcome) and 6 months (secondary outcome) after the visit. The risk behavior scores were constructed for each participant by adding all of the risk behaviors for which the tool included feedback (alcohol use, marijuana or other drug use, driving while intoxicated, tobacco use, depression, texting while driving, inconsistent seatbelt use, inconsistent helmet use, unprotected sexual activity, high sugary beverage intake, low fruit and vegetable intake, inadequate sleep, and low physical activity) at 3 and 6 months. High-risk behaviors were assigned a score of 2, moderate-risk behaviors were assigned a score of 1, and low-risk behaviors were assigned a score of 0 (score potential range: 0-21, further details regarding items in Multimedia Appendix 2). Treating baseline, 3-month, and 6-month risk scores as repeated measures, we applied linear mixed effects regression models to compare changes over time in adolescent-reported total risk scores at 3 and 6 months, relative to baseline, in intervention versus control adolescents controlling for baseline sex, age, and clinic as a random effect. To examine the effects of the intervention on health risk behaviors, we conducted exploratory logistic regression analyses for individual risk behaviors. Owing to concerns about estimate instability, we did not conduct analyses for individual behaviors in which fewer than 10 adolescents per study arm endorsed the behavior.

Patient Satisfaction Outcome

Patient satisfaction was measured on the 1-day postvisit survey using a satisfaction scale ranging from 1 to 10 from the Consumer Assessment of Health care Providers and Systems [28].

Differences between groups were examined using linear mixed effect regression while controlling for baseline age, sex, and clinic-specific random effects.

Control adolescents were the reference group for all regression analyses. For mixed effects Poisson regression, we determined that an estimate was statistically significant if its 95% CI for the rate ratio did not include 1. For the mixed effects linear regression models, statistical significance was based on *P* values calculated using the Satterwaite degrees of freedom method [29].

Results

Overview

In total, letters were sent to 1665 homes inviting adolescents to participate (Multimedia Appendix 4). The final study sample that completed all consent and baseline procedures was 301 adolescents (301/1586, 18.9% of the eligible sample). One adolescent withdrew from the study and requested that their data not be used, leaving an analytic sample of 300 adolescents. After consent, 145 patients were randomized to the intervention group and 155 to the control group. The response rates at 1 day, 3 months, and 6 months were 94% (282/300), 94.3% (283/300), and 94.6% (284/300), respectively.

Baseline Demographics and Risk Assessment

Randomization was balanced, with no differences between intervention and control adolescents in terms of demographics or baseline risk score (Table 1). Among the 300 participants, 43% (n=129) were female, 76% (n=228) were between the ages of 13 and 15 years, and 24% (n=72) were aged 16-18 years. Most adolescents identified as White (n=192, 64%), with the next largest group identifying as being of more than one race or “other” (n=55, 18.3%). In total, 92% (n=276) of adolescents had at least one health risk behavior at baseline, with a mean baseline risk score of 2.86 (SD 2.33) for intervention and 3.10 (SD 2.52) for control participants.

Table 2 summarizes the reported risk behaviors in order of baseline frequency at baseline, 3 months, and 6 months, with the most common risk behavior being low fruit and vegetable intake and the least frequent being driving under influence.

Table 1. Sample demographics.

Characteristic	Control (n=155)	Intervention (n=145)
Gender, n (%)		
Female	70 (45.2)	59 (40.7)
Male	82 (52.9)	86 (59.3)
Trans or nonbinary	3 (1.9)	0 (0.0)
Age (years), n (%)		
13-15	114 (73.5)	114 (78.6)
16-18	41 (26.4)	31 (21.3)
Race or ethnicity, n (%)		
White	99 (63.9)	93 (64.1)
Hispanic	12 (7.7)	7 (4.8)
African American	13 (8.4)	6 (4.2)
Asian or Pacific Islander	7 (4.5)	7 (4.8)
Native American	0 (0.0)	1 (0.7)
Other or more than one	24 (15.5)	31 (21.4)
Risk behavior score at baseline, mean (SD)	3.10 (2.52)	2.86 (2.33)

Table 2. Prevalence of individual risk behaviors over time in intervention and control adolescents.

Behavior	Intervention			Control			Logistic regression, *P* value[a]
	Baseline (n=145), n (%)	3 months (n=138), n (%)	6 months (n=139), n (%)	Baseline (n=155), n (%)	3 months (n=145), n (%)	6 months (n=145), n (%)	
Low fruit or vegetable intake	115 (79.3)	106 (76.8)	98 (70.5)	132 (85.1)	118 (81.4)	111 (76.6)	.93
Low sleep time	46 (31.7)	52 (37.7)	65 (46.8)	54 (34.8)	55 (37.9)	72 (49.6)	.89
Low physical activity	39 (26.9)	44 (31.9)	36 (25.9)	50 (32.3)	52 (35.9)	51 (35.2)	.77
Inconsistent helmet use	37 (25.5)	24 (17.4)	22 (15.8)	39 (25.2)	25 (17.2)	20 (13.8)	.54
High sugary beverage intake	28 (19.3)	39 (28.2)	36 (25.9)	36 (23.2)	37 (25.5)	35 (24.1)	.47
Depression	13 (9.0)	15 (10.9)	14 (10.1)	23 (14.8)	15 (10.3)	18 (12.4)	.24
Inconsistent seatbelt use	16 (11.0)	7 (5.1)	11 (7.9)	10 (6.5)	7 (4.8)	10 (6.9)	.11
Texting while driving	9 (5.8)	10 (7.2)	8 (5.8)	13 (8.4)	8 (5.5)	10 (6.9)	—[b]
Marijuana use	10 (6.9)	4 (2.9)	3 (2.2)	7 (4.5)	4 (2.8)	5 (3.4)	—
Alcohol use	8 (5.5)	3 (2.2)	4 (2.9)	6 (3.9)	4 (2.8)	4 (2.8)	—
Tobacco use	4 (2.8)	0 (0.0)	1 (0.7)	4 (2.6)	2 (1.4)	3 (2.1)	—
Sexual risk	1 (0.7)	3 (2.2)	3 (2.2)	4 (2.6)	2 (1.4)	3 (2.1)	—
Driving under the influence	0 (0.0)	0 (0.0)	0 (0.0)	2 (1.3)	1 (0.7)	0 (0.0)	—

[a]*P* values were based on the likelihood ratio test comparing mixed effects logistic regression with and without period-by-group interaction. Both models controlled for random effects corresponding to within-individual clustering.

[b]Owing to concerns about estimate instability, we did not conduct analyses for individual behaviors in which fewer than 10 adolescents per study arm endorsed the behavior.

Clinician Counseling Analysis Results

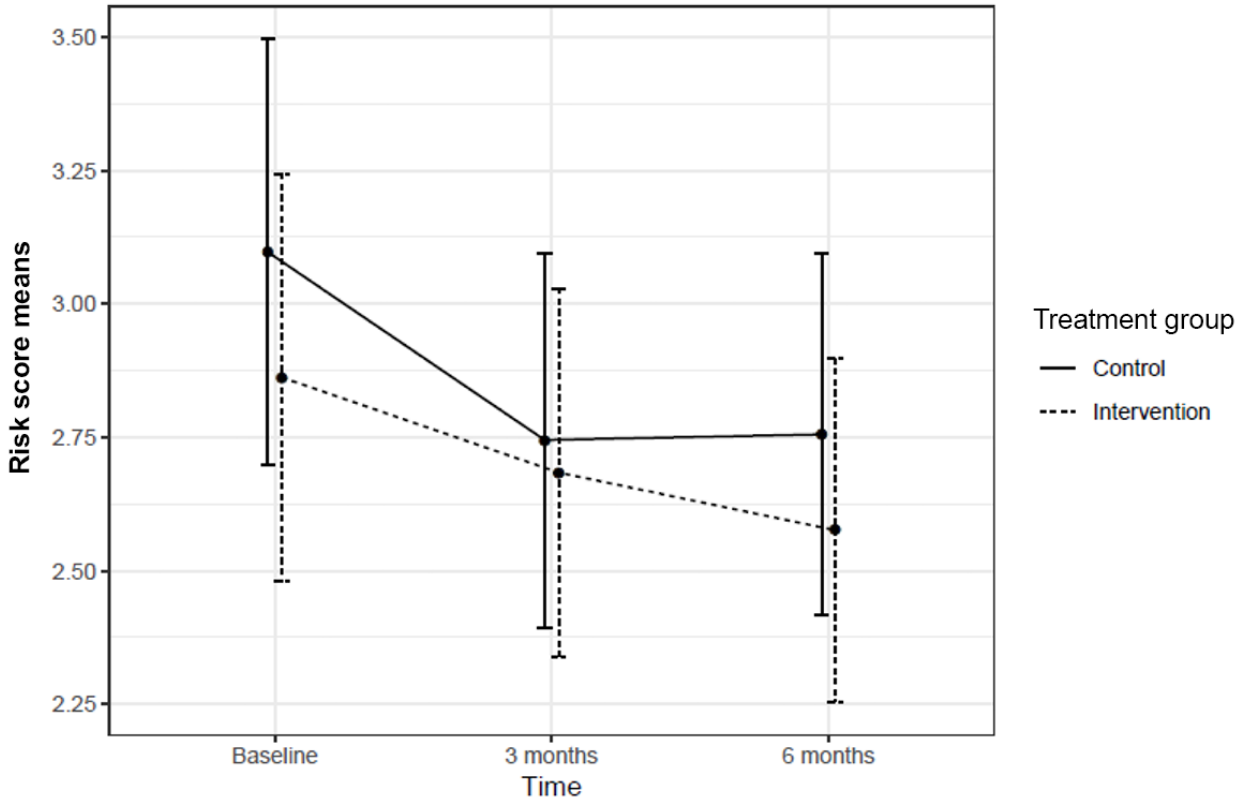
Among control adolescents, 380 moderate- and high-risk behaviors were endorsed, among which adolescents reported receiving clinician counseling for 148 (38.9%) behaviors during the visit. Intervention adolescents reported a total of 326 moderate- and high-risk behaviors, among which 184 (56.4%) were counseled by clinicians during the visits. In the Poisson regression analyses, intervention adolescents were significantly more likely to report that they had received counseling for their endorsed moderate- and high-risk behaviors than control adolescents (adjusted rate ratio [aRR] 1.36, 95% CI 1.04 to 1.78). To examine the impact of the intervention on rates of counseling by risk behavior severity level, we also examined rates of counseling for intervention and control adolescents based on whether they were classified as low, moderate, or high risk. Intervention adolescents were 40% more likely than adolescents in the control group to have received counseling for moderate-risk behaviors (aRR 1.40, 95% CI 1.09 to 1.80). For high-risk behaviors, the rate of counseling was 70% higher among intervention than control adolescents (aRR 1.70, 95% CI 1.06 to 2.74). There were no significant differences between intervention and control adolescents in reported counseling for no-/low-risk behaviors (aRR 1.12, 95% CI 0.85 to 1.48).

Risk Behavior and Patient Satisfaction Analyses

The baseline risk score was 2.86 (SD 2.33) for adolescents in the intervention group and 3.10 (SD 2.52) for adolescents in the control group, respectively (*P*=.40). At 3 months, the risk score for adolescents in the intervention group was 2.68 (SD 2.04) compared with 2.74 (SD 2.11), respectively, for adolescents in the control group (*P*=.81). At 6 months, the risk score for adolescents in the intervention group was 2.58 (SD 1.87) compared with 2.76 (SD 2.05) for adolescents in the control group (*P*=.45). In mixed effects linear regression analysis including both 3- and 6-month outcomes, there was a significant reduction in risk behaviors in both groups at 3 months (β=−.33; 95% CI −0.62 to −0.05; *P*=.02) and 6 months (β=−.29; 95% CI −0.57 to −0.01; *P*=.05). There were no significant differences in risk scores between the intervention and control groups at either time point (Figure 2). At 3 months, the score difference between groups was 0.15 (β=−.15; 95% CI −0.25 to 0.55; *P*=.47), and at 6 months, it was 0.12 (β=−.12; 95% CI −0.29 to 0.52; *P*=.57). In secondary analyses examining individual behaviors, no significant differences in the reduction of behaviors were observed between the adolescents of the intervention and control groups (Figure 2). There were also no significant differences between groups in patient satisfaction with the well-care visit process based on regression analysis controlling for age, gender, and clinic as a random effect

(intervention mean: 9.46, SD 0.79; control mean 9.27, SD 0.86; *P*=.07).

Figure 2. Health risk behavior scores in adolescents in the intervention and control groups by time.



Discussion

Principal Findings

In this study of an integrated screening and feedback tool, Check Yourself version 2, we found that adolescents in the intervention group were significantly more likely to report having been counseled by clinicians on risk behaviors than adolescents in the control group. However, despite significant differences in reported counseling between adolescents in the intervention and control groups, both groups demonstrated reductions in risk behavior scores, and there were no significant differences between the groups at 3 or 6 months after the intervention. There were also no significant differences in satisfaction between the 2 groups. These results are in contrast to our original study [21], which showed both an increase in reported counseling and a reduction in risk behavior scores at 3 months for youth in the intervention group as compared with controls. This study further adds to the growing body of literature on multi-behavior screening and preventive counseling interventions in adolescent well-care visits, which suggests that although provider counseling can be increased, the effects on risk behavior reductions are modest and inconsistent across studies [7].

In comparing the results to our prior study, it is important to note that this study tested a modified version of the tool with increased interactive content, which allowed adolescents to control the amount of information they received. Although adolescents requested incorporating these choices in content viewing, it is possible that the new adaptations resulted in less content exposure, particularly for at-risk adolescents who were not concerned about their behaviors. In this study design, we did not have the ability to assess how long adolescents spent on specific components of the feedback, although we do know that the overall time spent in this version of the tool was longer than the prior tool version. Future studies of interactive eHealth tools such as this could provide a better understanding of how risk influences engagement in feedback content.

It is also important to note that both the intervention and control groups experienced decreases in risk 3 months following their well-care visit in this sample. The reduction in risk in the control group would have weakened the ability to detect a difference. At baseline, all of the clinics in the study indicated that they conducted some form of paper and interview assessment of risk behaviors during the well-care visit. We collected information on the paper tools implemented and did not find substantial overlap with the risk behavior screening content of the electronic tools; however, all control teens completed an electronic health risk behavior assessment as part of the baseline study procedures. Although the results of this screening were not provided to the clinicians, it is possible that even in the absence of feedback, participating in the electronic screening may have resulted in behavior change, as teens reflected on their responses to risk behavior questions. In addition, as participants were randomized at the individual level, it is possible that some of the improvement in the control group was due to spillover effects as study clinicians and clinic staff applied learning from working with adolescents in the intervention group to control group. Our clinician counseling measure was based on

adolescent self-report and did not allow us to directly assess the content of counseling delivered during the visits to test this possibility.

Unlike our prior study, we also collected 6-month outcomes that allowed us to examine the long-term effects of the intervention. Although it was encouraging to see that risk scores continued to trend downward for the intervention sample, the differences between the control and intervention groups were not significant. Given the lack of effect at 3 months, it is difficult to draw conclusions from the 6-month data. Furthermore, 2 prior studies that examined both short-term (3 months) and long-term outcomes (12 months) found that significant differences in risk behaviors noted at 3 months were no longer significant at 12 months [30,31]. These 2 studies employed different models of brief interventions. One involved 9 hours of clinician training in motivational interviewing and system support for the implementation of a screening tool for all visits among eligible adolescents and young adults [30]. The second intervention focused on those aged 14 and 15 years enrolled in 8 general practice sites who were invited to participate in a 20-minute health consultation on risk behaviors of their choosing with a trained nurse [31]. Other studies that have examined 6- or 12-month outcomes have found significant improvements in single outcomes—helmet use [32,33] and exercise [19] only. Given the health care resources directed at screening and preventive counseling, understanding the long-term impacts of multiple risk behavior interventions is an area worthy of future study.

Limitations

This study had several limitations. First, although the use of a combined risk behavior outcome measure allowed us to test across the full range of behaviors for which clinicians were providing counseling, it is more difficult to interpret. We selected this measure as we feel it is more consistent with the multi-risk focus of behavioral counseling delivered in the pediatric well-care visit setting. However, this approach limits the conclusions we can draw regarding changes in any specific behavior. We conducted secondary analyses of individual behaviors to allow for a more ready interpretation of the intervention; however, for many behaviors, the prevalence at baseline was too low to draw conclusions on behavior change. The use of this multi-risk measure also limits our ability to compare outcomes with other studies, as prior research has measured a range of individual behavior outcomes [7].

A second limitation of this study is the low prevalence of individual behaviors. Consistent with other studies in pediatric primary care [34,35], including our own [21], adolescents receiving well-care tended to be younger: 76% (228/300) of participants were in the 13- to 15-year-old age group. Younger adolescents are less likely to engage in risk behaviors than older adolescents, which may limit their ability to show changes in behaviors. It is also possible that adolescents are less likely to endorse risk in the setting of a well-child visit because of concerns about confidentiality. Our research with this tool in a school-based clinic setting demonstrated significantly higher rates of youth-reported risk behaviors even after matching for age [36]. Prior research has also suggested that acute visits may be a more effective platform for risk screening among adolescents [37,38]. To increase the effective delivery of counseling, more research is needed to identify the best venues for reaching older and at-risk adolescents, including the added benefits and costs of screening at acute visits as well as screening in school-based health settings. Finally, this study was conducted among adolescents who visited primary care clinics in the Pacific Northwest and may not be generalizable to other settings.

Conclusions

Despite these limitations, this study adds to the literature regarding the use of eHealth tools in screening and preventive care for adolescents and raises important questions worthy of further study. Health risk behaviors have a significant influence on morbidity and mortality during adolescence and adulthood and guidelines recommend screening and intervention during adolescent well-care visits. Electronic screening has been repeatedly shown to increase provider identification of risk. This study further demonstrates that the addition of feedback for adolescents and results for clinicians increases clinician counseling. Electronic platforms such as these can be important tools for future research to examine the impact of components and types of preventive content to effect behavior change as well as how to reach the adolescents who would most benefit.

Acknowledgments

This research was primarily supported by the Agency for Healthcare Research and Quality (AHRQ 5R01HS023383; principal investigator: CM). The granting agency that supported this research approved the study design and received periodic updates on data collection but was not involved in the analysis of the data, the decision to submit a manuscript, or in the writing of the manuscript itself. Additional support was provided by the Health Resources and Services Administration of the US Department of Health and Human Services under cooperative agreement UA6MC27378, Adolescent and Young Adult Health Research Network. LR and CM designed the study. EP and CZ conducted data analysis. EO and JK advised on the study design and reporting of results. LR and CM were responsible for the development of the feedback content in the Check Yourself version 2.0 tool with input from JK. The tool is licensed to Tickit Health [39].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Check Yourself (version 2) screenshots.
[PDF File (Adobe PDF File), 771 KB - jmir_v23i3e24135_app1.pdf]

Multimedia Appendix 2
Risk behaviors included in overall summary outcome measure.
[PDF File (Adobe PDF File), 83 KB - jmir_v23i3e24135_app2.pdf]

Multimedia Appendix 3
Check Yourself app overview.
[PDF File (Adobe PDF File), 227 KB - jmir_v23i3e24135_app3.pdf]

Multimedia Appendix 4
CONSORT (Consolidated Standards of Reporting Trials) diagram.
[PDF File (Adobe PDF File), 18 KB - jmir_v23i3e24135_app4.pdf]

References

1. McGinnis JM, Foege WH. Actual causes of death in the United States. J Am Med Assoc 1993 Nov 10;270(18):2207-2212. [Medline: 8411605]
2. Park MJ, Scott JT, Adams SH, Brindis CD, Irwin CE. Adolescent and young adult health in the United States in the past decade: little improvement and young adults remain worse off than adolescents. J Adolesc Health 2014 Jul;55(1):3-16. [doi: 10.1016/j.jadohealth.2014.04.003] [Medline: 24815958]
3. Hagan JF, Shaw JS, Duncan PM. Bright futures: guidelines for health supervision of infants, children, and adolescents. 4th edition. Elk Grove Village, IL: American Academy of Pediatrics; 2017.
4. Elster A, Kuznets N. Guidelines for Adolescent Preventive Services (GAPS). Baltimore, MD: Wiliams & Wilkins; 1994. URL: https://www.uptodate.com/contents/guidelines-for-adolescent-preventive-services [accessed 2021-02-25]
5. Halpern-Felsher BL, Ozer EM, Millstein SG, Wibbelsman CJ, Fuster CD, Elster AB, et al. Preventive services in a health maintenance organization: how well do pediatricians screen and educate adolescent patients? Arch Pediatr Adolesc Med 2000 Feb;154(2):173-179. [doi: 10.1001/archpedi.154.2.173] [Medline: 10665605]
6. Olson AL, Gaffney CA, Hedberg VA, Gladstone GR. Use of inexpensive technology to enhance adolescent health screening and counseling. Arch Pediatr Adolesc Med 2009 Feb;163(2):172-177. [doi: 10.1001/archpediatrics.2008.533] [Medline: 19188650]
7. Webb MJ, Kauer SD, Ozer EM, Haller DM, Sanci LA. Does screening for and intervening with multiple health compromising behaviours and mental health disorders amongst young people attending primary care improve health outcomes? A systematic review. BMC Fam Pract 2016 Aug 04;17:104 [FREE Full text] [doi: 10.1186/s12875-016-0504-1] [Medline: 27488823]
8. Stevens J, Kelleher KJ, Gardner W, Chisolm D, McGeehan J, Pajer K, et al. Trial of computerized screening for adolescent behavioral concerns. Pediatrics 2008 Jun;121(6):1099-1105. [doi: 10.1542/peds.2007-1878] [Medline: 18519478]
9. Harris SK, Aalsma MC, Weitzman ER, Garcia-Huidobro D, Wong C, Hadland SE, et al. Research on clinical preventive services for adolescents and young adults: where are we and where do we need to go? J Adolesc Health 2017 Mar;60(3):249-260 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.10.005] [Medline: 28011064]
10. Williams ML, Freeman RC, Bowen AM, Zhao Z, Elwood WN, Gordon C, et al. A comparison of the reliability of self-reported drug use and sexual behaviors using computer-assisted versus face-to-face interviewing. AIDS Educ Prev 2000 Jun;12(3):199-213. [Medline: 10926124]
11. Murphy DA, Durako S, Muenz LR, Wilson CM. Marijuana use among HIV-positive and high-risk adolescents: a comparison of self-report through audio computer-assisted self-administered interviewing and urinalysis. Am J Epidemiol 2000 Nov 01;152(9):805-813. [doi: 10.1093/aje/152.9.805] [Medline: 11085391]
12. Suler J. The online disinhibition effect. Cyberpsychol Behav 2004 Jun;7(3):321-326. [doi: 10.1089/1094931041291295] [Medline: 15257832]
13. Jasik CB, Berna M, Martin M, Ozer EM. Teen preferences for clinic-based behavior screens: who, where, when, and how? J Adolesc Health 2016 Dec;59(6):722-724. [doi: 10.1016/j.jadohealth.2016.08.009] [Medline: 27884300]
14. Sterling S, Kline-Simon AH, Wibbelsman C, Wong A, Weisner C. Screening for adolescent alcohol and drug use in pediatric health-care settings: predictors and implications for practice and policy. Addict Sci Clin Pract 2012 Aug 16;7:13 [FREE Full text] [doi: 10.1186/1940-0640-7-13] [Medline: 23186254]
15. Paperny DM. Computerized health assessment and education for adolescent HIV and STD prevention in health care settings and schools. Health Educ Behav 1997 Feb;24(1):54-70. [doi: 10.1177/109019819702400107] [Medline: 9112098]
16. Paperny DM, Hedberg VA. Computer-assisted health counselor visits: a low-cost model for comprehensive adolescent preventive services. Arch Pediatr Adolesc Med 1999 Jan;153(1):63-67. [doi: 10.1001/archpedi.153.1.63] [Medline: 9895001]

17. Paperny DM, Aono JY, Lehman RM, Hammar SL, Risser J. Computer-assisted detection and intervention in adolescent high-risk health behaviors. The Journal of Pediatrics 1990 Mar;116(3):456-462 [FREE Full text] [doi: 10.1016/s0022-3476(05)82844-6]
18. Coker TR, Sareen HG, Chung PJ, Kennedy DP, Weidmer BA, Schuster MA. Improving access to and utilization of adolescent preventive health care: the perspectives of adolescents and parents. J Adolesc Health 2010 Aug;47(2):133-142. [doi: 10.1016/j.jadohealth.2010.01.005] [Medline: 20638005]
19. Olson AL, Gaffney CA, Lee PW, Starr P. Changing adolescent health behaviors: the healthy teens counseling approach. Am J Prev Med 2008 Nov;35(5 Suppl):S359-S364. [doi: 10.1016/j.amepre.2008.08.014] [Medline: 18929982]
20. Webb MJ, Wadley G, Sanci LA. Improving patient-centered care for young people in general practice with a codesigned screening app: mixed methods study. JMIR Mhealth Uhealth 2017 Aug 11;5(8):e118 [FREE Full text] [doi: 10.2196/mhealth.7816] [Medline: 28801302]
21. Richardson LP, Zhou C, Gersh E, Spielvogle H, Taylor JA, McCarty CA. Effect of electronic screening with personalized feedback on adolescent health risk behaviors in a primary care setting: a randomized clinical trial. JAMA Netw Open 2019 May 03;2(5):e193581 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.3581] [Medline: 31074815]
22. Maguire M. Methods to support human-centred design. Int J of Human-Comput Stud. URL: https://doi.org/10.1006/ijhc.2001.0503 [accessed 2021-02-05]
23. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
24. Goldenring JM, Cohen E. Getting into adolescent heads. Contemp Pediatr. 1988. URL: https://www.contemporarypediatrics.com/view/getting-adolescent-heads [accessed 2021-02-25]
25. Zieve GG, Richardson LP, Katzman K, Spielvogle H, Whitehouse S, McCarty CA. Adolescents' perspectives on personalized e-feedback in the context of health risk behavior screening for primary care: qualitative study. J Med Internet Res 2017 Jul 20;19(7):e261 [FREE Full text] [doi: 10.2196/jmir.7474] [Medline: 28729236]
26. Ozer EM, Adams SH, Lustig JL, Millstein SG, Camfield K, El-Diwany S, et al. Can it be done? Implementing adolescent clinical preventive services. Health Serv Res 2001 Dec;36(6 Pt 2):150-165 [FREE Full text] [Medline: 16148966]
27. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Project for Statistical Computing; 2018. URL: https://R-project.org/ [accessed 2021-02-25]
28. Darby C, Crofton C, Clancy CM. Consumer Assessment of Health Providers and Systems (CAHPS): evolving to meet stakeholder needs. Am J Med Qual 2006;21(2):144-147. [doi: 10.1177/1062860606286263] [Medline: 16533906]
29. Kuznetsova ABP, Brockhoff PB, Christensen RHB. Package: tests in linear mixed effects models. J Stat Soft 2017;82(13):1-26. [doi: 10.18637/jss.v082.i13]
30. Sanci L, Chondros P, Sawyer S, Pirkis J, Ozer E, Hegarty K, et al. Responding to young people's health risks in primary care: a cluster randomised trial of training clinicians in screening and motivational interviewing. PLoS One 2015;10(9):- [FREE Full text] [doi: 10.1371/journal.pone.0137581] [Medline: 26422235]
31. Walker Z, Townsend J, Oakley L, Donovan C, Smith H, Hurst Z, et al. Health promotion for adolescents in primary care: randomised controlled trial. Br Med J 2002 Sep 07;325(7363):524 [FREE Full text] [doi: 10.1136/bmj.325.7363.524] [Medline: 12217993]
32. Ozer EM, Adams SH, Orrell-Valente JK, Wibbelsman CJ, Lustig JL, Millstein SG, et al. Does delivering preventive services in primary care reduce adolescent risky behavior? J Adolesc Health 2011 Nov;49(5):476-482. [doi: 10.1016/j.jadohealth.2011.02.011] [Medline: 22018561]
33. Stevens MM, Olson AL, Gaffney CA, Tosteson TD, Mott LA, Starr P. A pediatric, practice-based, randomized trial of drinking and smoking prevention and bicycle helmet, gun, and seatbelt safety promotion. Pediatrics 2002 Mar;109(3):490-497. [doi: 10.1542/peds.109.3.490] [Medline: 11875146]
34. Anand V, Carroll AE, Downs SM. Automated primary care screening in pediatric waiting rooms. Pediatrics 2012 May;129(5):e1275-e1281 [FREE Full text] [doi: 10.1542/peds.2011-2875] [Medline: 22508925]
35. Rand CM, Shone LP, Albertin C, Auinger P, Klein JD, Szilagyi PG. National health care visit patterns of adolescents: implications for delivery of new adolescent vaccines. Arch Pediatr Adolesc Med 2007 Mar;161(3):252-259. [doi: 10.1001/archpedi.161.3.252] [Medline: 17339506]
36. McCarty CA, Gersh E, Katzman K, Lee CM, Sucato GS, Richardson LP. Screening and brief intervention with adolescents with risky alcohol use in school-based health centers: a randomized clinical trial of the tool. Subst Abus 2019;40(4):510-518. [doi: 10.1080/08897077.2019.1576090] [Medline: 30883284]
37. Borowsky IW, Mozayeny S, Ireland M. Brief psychosocial screening at health supervision and acute care visits. Pediatrics 2003 Jul;112(1 Pt 1):129-133. [doi: 10.1542/peds.112.1.129] [Medline: 12837878]
38. Hacker KA, Arsenault LN, Williams S, Digirolamo AM. Mental and behavioral health screening at preventive visits: opportunities for follow-up of patients who are nonadherent with the next preventive visit. J Pediatr 2011 Apr;158(4):666-671. [doi: 10.1016/j.jpeds.2010.09.059] [Medline: 21074180]
39. Tickit Health. URL: https://tickithealth.com/ [accessed 2021-03-06]

Abbreviations

aRR: adjusted rate ratio
MICE: multiple imputation with chained equations

Edited by G Eysenbach; submitted 04.09.20; peer-reviewed by A Radovic-Stakic, K Kelleher; comments to author 26.09.20; revised version received 30.12.20; accepted 16.01.21; published 12.03.21.

Please cite as:
Richardson L, Parker EO, Zhou C, Kientz J, Ozer E, McCarty C
Electronic Health Risk Behavior Screening With Integrated Feedback Among Adolescents in Primary Care: Randomized Controlled Trial
J Med Internet Res 2021;23(3):e24135
URL: https://www.jmir.org/2021/3/e24135
doi:10.2196/24135
PMID:33709942

Original Paper

Risk Factors and Leadership in a Digitalized Working World and Their Effects on Employees' Stress and Resources: Web-Based Questionnaire Study

Anita Bregenzer[1], PhD; Paulino Jimenez[1], PhD

Institute of Psychology, University of Graz, Graz, Austria

Corresponding Author:
Paulino Jimenez, PhD
Institute of Psychology
University of Graz
Universitätsplatz 2
Graz, 8010
Austria
Phone: 43 316380 ext 5128
Email: paul.jimenez@uni-graz.at

Abstract

Background: In today's world of work, the digitalization of work and communication processes is increasing, and will increase even further. This increase in digitalization at the workplace brings many new aspects of working life to light, such as working in virtual teams, mobile working, expectations of being constantly available, and the need for support in adapting and learning new digital tools. These changes to the workplace can contain risks that might harm the well-being of employees. Leaders can support the well-being of their employees in terms of protecting and replenishing their work-related resources to cope with critical work demands. This so-called health-promoting leadership could serve as a buffer between risk at the workplace and critical outcomes, such as stress, by amplifying work-related resources.

Objective: This study's aims were twofold. First, we wanted to investigate if risk factors related to higher digitalization at the workplace can be identified and if these risk factors have an impairing effect on the well-being of employees (eg, higher stress and lower resources). Second, we wanted to investigate if the health-impairing effects of these risk factors can be reduced by health-promoting leadership.

Methods: A total of 1412 employees from Austria, Germany, and Switzerland took part in this online study and provided information on their perceived risks at the workplace, their leaders' health-promoting behaviors, and their work-related stress and resources.

Results: The results of a hierarchical regression analysis showed that all four risk factors of digital work (distributed team work, mobile work, constant availability, and inefficient technical support) were related to higher stress at the workplace. In addition, distributed team work and inefficient technical support were associated with lower work-related resources. A possible buffer effect of health-promoting leadership between these risks and employee well-being was visible for inefficient technical support. In particular, in the case of having fewer support opportunities in learning and using digital tools, leaders could weaken the potential critical effects on stress. As for the other risk factors, leaders might engage in a different leadership behavior to improve their employees' well-being, as the physical distance between leaders and employees in virtual team work or mobile work could make health-promoting leadership more difficult.

Conclusions: In a digitalized working world, solutions are needed to create working conditions that benefit employees. The results of this study strongly support the importance of investigating risk factors associated with an increase in digitalization at the workplace in addition to traditional risk factors. As for leadership, leaders need to show leadership behavior adapted to a digitalized workplace in order to reduce employee stress and increase work-related resources.

(J Med Internet Res 2021;23(3):e24906) doi:10.2196/24906

KEYWORDS

digitalization; leadership; new ways of working; resources; stress

Introduction

Background

In the past years, the digitalization of the workplace has been studied more as a phenomenon relevant to a small number of people than as an important and necessary step to improve the working world. Digitalization is currently affecting many areas and will continue to do so in the future, so the effects of digitalization in the workplace must be studied more closely in relation to work processes within the company and in relation to the well-being and performance of employees. Digitalized work brings many new aspects of working life to light, such as working in virtual teams, mobile working, blur between leisure and work, expectations of constant availability, and the frequent need to adapt to digital changes and learn new digital tools [1,2]. Organizations must be able to react adequately to these changes in order to minimize possible critical effects at individual and team levels (eg, stress, engagement, and performance). Owing to the speed at which digitalization is entering the current world of work, solutions are needed just as quickly, as organizations then can prepare their employees optimally for the newly emerging forms of work.

The topic of digitalization of the working world is currently experiencing an upswing in scientific research, especially under the term "new ways of working" (NWW). NWW describes changes to the workplace that take place in the following four aspects: physical workplace, information and communication technology (ICT), organization and management, and work culture [3]. For example, an important aspect of NWW is having more flexibility in deciding when and where employees can work, as well as using ICTs, such as email, smartphones, and videoconferences. It is expected that those aspects should lead to more efficient work processes [4]. Research in this area has focused strongly on the positive effects of new working forms on employees, such as higher engagement and performance [5,6]. However, there is evidence that these new forms of work also have critical effects on employees, such as fatigue and exhaustion [4,7].

Research in this area is important to highlight the risks of a nonoptimal design of a digitalized workplace. However, there is currently a lack of information on how the company and its employees can benefit optimally from increased digitalization of the workplace. One solution to improve working conditions for employees is leadership. Leaders can change their employees' working conditions and thus impact their employees' health by managing and allocating resources at the workplace [8]. More specifically, the concept of *health-promoting leadership* includes leadership behaviors that aim at providing resourceful working conditions for employees [9]. This in turn can reduce the negative consequences of critical working conditions such as stress [10]. Leaders can increase resources at the workplace, for example, by specifically supporting the community within the team or by giving their employees possibilities to participate in important decisions.

This study's aims were twofold. First, we wanted to investigate if risk factors related to a higher digitalization at the workplace can be identified. With the term "risk factors," we mainly followed the definition of mental risk factors (according to ISO 10075 [11]), which can have an impairing effect on the well-being of employees (eg, higher stress and lower resources). Second, we wanted to investigate if health-promoting leadership moderates the relationship between risk factors of digitalization at the workplace and employees' stress and resources. More specifically, it is of interest if the health-impairing effects of certain risk factors can be reduced by health-promoting leadership. To our knowledge, the role of health-promoting leadership in workplaces with increasing digitalization has not yet been addressed directly in research.

Digital Working World and Effects on Employee Well-Being

The working environment is an important context factor at the workplace that affects the health of employees. Being exposed to a critical working environment with high risks might result in negative psychological states that can negatively affect the individual's behavior at work [12]. These risks can be associated with the physical environment, the organizational and social environment, or the task itself [11,13]. The aim is to design the workplace in such a way that risk factors are minimized or at least the impairing health effect of these risk factors is reduced with specific interventions. However, the traditional working environment has changed through the application of ICT, and "new" working forms have emerged, such as virtual teams and mobile telework [14,15]. These changes are accompanied by new risks in the workplace, and the potential harmful effects on employee well-being have to be examined more closely.

Research in the field of NWW seems to highlight the positive aspects of a digitalized workplace. ten Brummelhuis et al [4] defined NWW as "…a work design in which employees can control the timing and place of their work, while being supported by electronic communication." Indeed, research shows that when employees experience more freedom in managing one's own time (ie, home office), the work is experienced as less stressful for employees [16]. However, NWW can also impact the employee's well-being negatively, such as having more blurred work-home boundaries, more fatigue, and higher mental demands [7]. Research also indicates that NWW might decrease resources such as autonomy. In a study conducted by van Steenbergen et al [16], employees worked in an organization where they could choose to work at home or at the office. However, the organization seemed to prefer work from home, and this preference for a home office might have been expressed by the company in such a strong way that employees experienced a lower feeling of autonomy.

These findings show that positive effects of NWW should not be expected automatically. On the contrary, NWW might include risks that could lead to harmful effects such as higher employee stress [7]. In a systematic review, the authors outlined the positive and negative aspects of NWW with ICT-enabled workers who were flexible in their work [7]. They found that factors, such as geographically distributed team work ("virtual teams"), time- and location-independent work ("mobile working"), and use of information technology at work, might have negative psychological impacts on the well-being of employees and should be addressed when NWW is

implemented. Distributed teamwork or mobile work and increased digital communication are also related to a feeling of having to be constantly available, which can also have a negative impact on well-being [1]. Other factors related to NWW, such as higher flexibility, access to organizational knowledge, and independent management of output, are mostly positive factors that benefit the well-being of employees [7]. However, research should focus more on risks to help organizations adequately address these risks at the workplace. The dimensions of NWW that include possible risks and their relationship with the stress of employees are described in detail below.

Geographically Distributed Team Work

Geographically distributed team work (referred to as "distributed team work" henceforth) has already been extensively investigated in the past under the term "virtual teams" [14]. In virtual teams, "…teams work together over time and distance via electronic media to combine effort and achieve common goals" [17]. Distributed team work has advantages as well as disadvantages. The advantages include reduced travel time and costs, being independent from time and place, including physically disadvantaged employees in the team, and working in a diverse heterogenous team [18]. However, the disadvantages have been studied in more detail. Owing to the geographical distance of team members, it is difficult to form group cohesion, which is why communication is less frequent and conflicts can occur more often than in face-to-face teams [19,20]. Furthermore, with the use of virtual media, important auditory and visual cues are not perceived sufficiently, which makes communication more difficult [21].

The critical effects of distributed team work on employees' stress have already been studied [22,23]. For example, virtual teams have more conflicts than traditional face-to-face teams and have difficulties in applying conflict management strategies [21]. More conflicts within the team result in more stress [24]. Stress can occur because of the excessive use of virtual communication media as well (eg, email flood [23]).

Mobile Work

The use of mobile devices allows employees to work in a distributed team and to work independent of time and location, because messages can now be sent and received from anywhere and at any time. Thus, high flexibility in the daily work routine can be achieved [25]. Work can be done from one's own home, from an external location, or from another continent. For this kind of work, the term "mobile telework" is commonly used in the literature, which is described as "work at a range of locations, spending regular and significant amount of time away from any office or home location" [26].

Mobile telework can differ from one job to another. There are jobs in which the place of work changes several times a week or day and employees cannot freely choose the work location, for example, work in sectors such as wholesale and retail trade, manufacturing, transportation and storage, information and communication, public administration, and health [27]. The effect of mobile telework on employee well-being can be quite different for those who have control over their working location as compared to that for those who have little say in where they must work [28].

Mobile telework is seen as a resource, especially when you can decide yourself where and when you work [29,30]. However, the physical distance between team members and leaders can reduce the quality of the relationship between employees and leaders [31]. Mobile telework is particularly demanding when the work location is uncertain or when the employees have less flexibility in organizing their work time [32]. Interruptions and distractions can occur more easily in mobile telework than in fixed workplaces (such as offices); for example, interruptions and distractions are more frequent in trains or in public places [32]. In addition, working at multiple locations increases mental demands, such as the feeling of "timeless" continuous work, constant changing of the rhythm of work, and reduced professional and social interaction [15].

Constant Availability

The use of ICT makes it easier to stay in touch with leaders, colleagues, customers, and family, as contact can now be made anywhere and at any time. This leads to the impression that people are available anywhere and at any time, which can have critical effects on employees' well-being [33,34].

The expectation of having to be constantly available for work leads to difficulties detaching from work during leisure time and to a stronger work-home interference [35,36]. Especially when the experienced work-home interference is high, using the smartphone for work-related purposes after work has a critical effect on employees' recovery process [37]. A longitudinal study showed that being constantly available for work increases emotional exhaustion over time [38]. Constantly receiving and checking work-related messages might lead to information overload, as people struggle with managing the inflow of messages [39]. This struggle to keep up with the increasing amount of information leads to higher stress [40,41].

On the contrary, being constantly available can have benefits for employees' well-being. In a study conducted by ten Brummelhuis et al [4], being constantly available through the use of mobile communication tools was associated with greater engagement. The authors argued that constant availability via email or telephone was associated with greater work flexibility, which was perceived as an advantage by employees. On the other hand, constant availability was also associated with more interruptions at work, which caused more exhaustion among employees.

Learning and Adapting to Digital Tools

The constant use of ICT for work-related activities raises another point that could be a risk factor for employees' well-being. In today's working world, new technologies are developed almost faster than people can learn and use them. The increasing amount and use of ICTs can lead to higher job demands in terms of mental and emotional overload, which might harm the well-being of employees [42]. Thus, the need for support in the use of digital tools and the need to build up competence in handling digital media are growing [43]. In the past, the increasing requirements to be able to handle digital tools were also investigated under the term "technostress." Technostress

is described as the mental stress that employees experience when they are asked to learn and use a new technology [44]. Weil and Rosen [44] found that technostress occurs if people are not taught how to handle technology adequately. Uncertainty about how to deal with new technology and the resulting inefficiency in dealing with modern technologies are currently still important issues in technostress research [45].

If the used technologies change too fast, employees experience difficulties in coping with the changes, which can raise work overload and stress [46]. On the contrary, adapting to new technologies at work might benefit the employees as well. Studies show that having a higher technological demand at work is related to engagement, indicating that learning new technological tools is perceived as challenging [47].

To ensure that technical changes in the workplace are experienced as positive challenges and not as hindrances, it is important that employees are adequately supported in learning and applying these technologies. For example, providing training or guidelines on how to deal with new media and having technical support at work are important for greater well-being at the workplace [48,49]. Social support from supervisors or colleagues is an important factor as well [42]. In the study by Knani et al [50], employees were introduced to a new technology at the workplace, which demanded high learning effort and led to higher emotional exhaustion. The critical effect on emotional exhaustion could reduce when employees experience high support from supervisors and employees. Atanasoff and Venable [51] added that employee-oriented leadership behavior is an important resource that might reduce the negative effects of digitalization, such as stress.

Digital Workplaces, Leadership, and Resources

Leaders in particular are challenged in a modern working environment. Research in the field of a home office and virtual teams has shown that leadership in a digitalized working environment has different requirements than in traditional work settings [52]. Working in a home office or virtual work in general requires a different role of leadership, in which the manager must lead strongly in an employee-oriented way [53]. An employee-oriented leadership is also preferred in working environments with high demands. According to Wegge et al [54], leader behavior can serve as a buffer between high work demands and critical outcomes, such as stress, by amplifying work-related resources at the workplace. Given the assumption that digitalized workplaces entail high demands, increasing work-related resources through leadership behavior is a particularly important aspect of supporting well-being in the workplace.

Work-related resources play a major role in the relationship between demands and stress [55]. Social resources (social support from colleagues) and task resources (autonomy, the possibility of participation, and the possibility of conducting breaks) are important work-related resources to reduce negative outcomes, such as stress and burnout. A highly digitalized workplace can contain risk factors that might lead to increased demands [46]. In workplaces with high demands, resources could be insufficiently gained, depleted, or even lost, which can cause stress and might increase the risk of getting burnout over time [56].

Maintaining and increasing work-related resources are therefore essential aspects of a health-promoting workplace. Leaders can support their employees in protecting and replenishing their work-related resources to cope with the demands of their work by showing health-promoting leadership behavior [57,58]. Health-promoting leadership is a positive leadership behavior, which enhances the work-related resources of employees. By changing working conditions (such as the health-promoting design of the six areas of work life [59]), it is possible to build up employees' work-related resources [60]. For example, leaders can ensure that work processes are organized in such a way that employees can cope well with increased workload. Leaders can give their employees opportunities to work autonomously and independently. Rewarding employees is also an essential aspect that can be undertaken by leaders in the form of positive feedback and appreciation. Leaders can strengthen the community in their team by encouraging open communication and mutual support. Acting fairly and paying attention to the values of employees are further aspects of health-promoting leadership [60].

Increasing work-related resources is also essential for a workplace with a high level of digitalization. Atanasoff and Venable [51] assumed that stress due to digitalization is related to lower work-related resources. According to the authors, important resources that should be increased are social support from colleagues, opportunities to participate in the use of technology, and clear information about technology. Therefore, health-promoting leadership could benefit a digitalized workplace as resources are preserved and restored.

The increasing digitalization of workplaces leads to changes in working conditions, which can be risk factors for reduced well-being and performance. Health-promoting leadership can minimize the negative effects of these risk factors by building up enough work-related resources to cope with these risk factors. This way of leadership behavior is described as the "buffer effect," which means leaders serve as a buffer against high work demands that might be a potential source for stress [54].

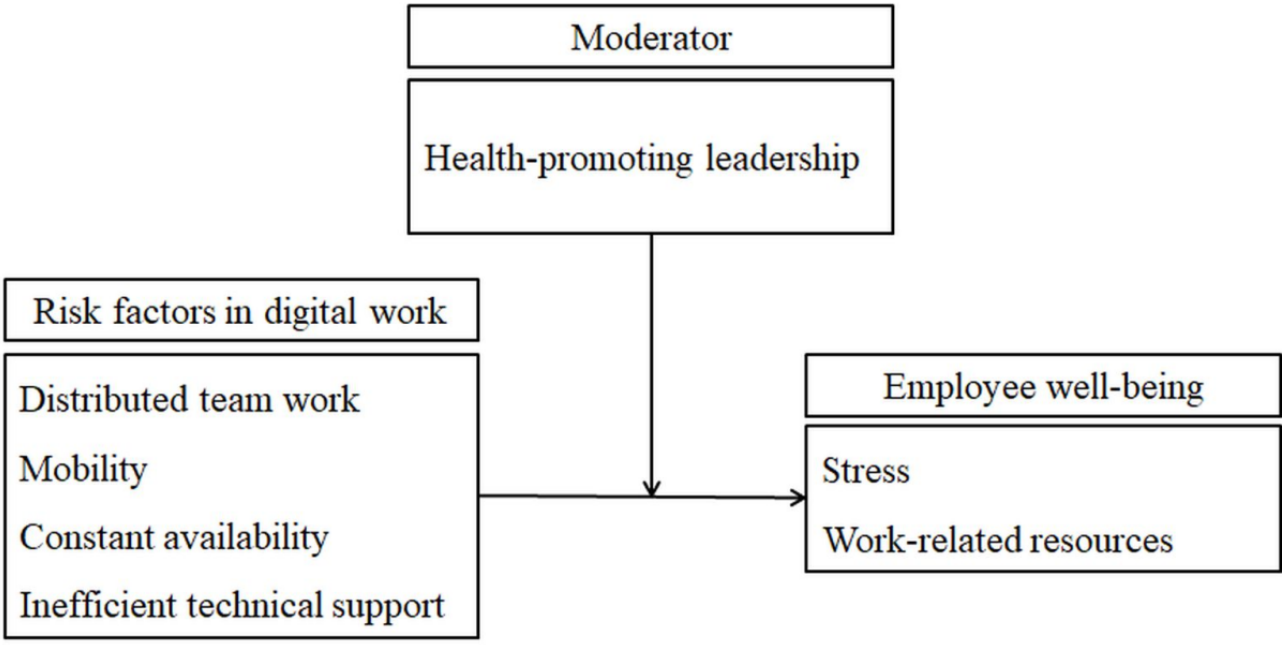
Study Aims and Hypotheses

In this study, we investigated the following four possible risk factors of digital work that could lead to higher stress and lower work-related resources among employees: distributed team work, mobile work, constant availability, and inefficient technical support. First, these four risk factors were examined with regard to their effects on the stress and work-related resources of employees. Second, a possible buffer effect of health-promoting leadership on the relationship of these risk factors with stress and work-related resources was analyzed. This will deepen the understanding of the importance of health-promoting leadership for digitalized workplaces and give an answer to the question of whether leadership behavior can reduce the potential harmful effects of risk factors in digitalized workplaces.

The following four hypotheses are proposed: (1) H1, risk factors in digital work (distributed team work, mobile work, constant

availability, and inefficient technical support) positively relate to employees' stress; (2) H2, risk factors in digital work (distributed team work, mobile work, constant availability, and inefficient technical support) negatively relate to employees' work-related resources; (3) H3, health-promoting leadership moderates the positive relationship between the risk factors in digital work and employees' stress (the relationship is weaker when health-promoting leadership is high); and (4) H4, health-promoting leadership moderates the negative relationship between the risk factors in digital work and employees' work-related resources (the relationship is weaker when health-promoting leadership is high). Figure 1 summarizes the overall conceptual model of the study.

Figure 1. Overall conceptual model of the study.



Methods

Participants and Procedures

The study was conducted as an online study with online questionnaires via the online survey platform Questback. The invitation for the online study was sent out in cooperation with a well-known German market research company. The participants of the study were recruited from the company's online panel. To obtain a heterogenous sample, we set the target male-to-female ratio at about 50:50. The same ratio was used for age (50% for <40 years and 50% for ≥40 years). As the online survey was in German, only German-speaking people were considered for recruitment (ie, persons from Germany, Austria, and Switzerland). The market research company contacted people in the online panel according to these specifications via email. The only criterion for participation in the study was work for at least 10 hours per week. If individuals stated in the questionnaire that they were working less than 10 hours per week, they were filtered out, and on the next page of the questionnaire, they were told that unfortunately they did not belong to the target group. The survey was then closed for this group.

On the first page of the survey, participants were informed about the purpose of the study, the length of the study, and the contact address of the research group. Participation was voluntary, and a small incentive was offered to people who completed all online questions.

Through this procedure, a representative sample of 1412 German-speaking workers in Austria (n=481, 34.06%), Germany (n=720, 50.99%), and Switzerland (n=211, 14.94%), who filled in all online questionnaires, was obtained. In this sample, 56.94% (804/1412) were women and 43.06% (608/1412) were men. The mean age was 41 years (mean 40.77 years, SD 12.30 years). Additionally, 24.36% (344/1412) had a graduate degree. On average, the participants worked 35 hours per week (mean 35.07 hours, SD 11.58 hours).

The participants worked in different business sectors. The majority worked in the service sector (257/1412, 18.20%), followed by health care (192/1412, 13.60%), commerce (167/1412, 11.83%), manufacturing (136/1412, 9.63%), and the public sector (127/1412, 8.99%).

Measures

Risks in Digital Work Scale

In the risks in digital work scale, 10 items measure different work characteristics in a digitalized workplace that could increase demands at the workplace for the following: (1) distributed team work, (2) mobile work, (3) constant availability, and (4) inefficient technical support. The items are written as statements and refer to the last 4 weeks ("How many times have you experienced the following aspects in the last 4 weeks?"). The 7-point scale ranges from 0 (never) to 6 (always). Example items for the four dimensions are shown in Table 1.

Table 1. Example items for the risks in digital work scale.

Construct/scale	Sample item
Distributed team work	My colleagues at other locations and I support each other (reversed).
Mobile work	Within a day, my work location changed.
Constant availability	I was available for work in my free time (eg, by telephone or email).
Inefficient technical support	I received support in case of uncertainties in the technical operation of devices, software, and others (reversed).

Health-Promoting Leadership

Health-promoting leadership was measured with the health-promoting leadership conditions questionnaire (HPLC) [9], where employees are able to evaluate the frequency of health-promoting leadership from their direct supervisor during the last 4 weeks. In this study, a short version with seven items was used, where each item can be related to one of the following seven aspects of health-promoting leadership: health awareness, workload, control, reward, community, fairness, and value fit. The items are rated on a 7-point scale ranging from 0 (never) to 6 (always). One example item for the dimension community is "In the last 4 weeks, my leader took care that…work is appreciated."

Stress and Resources

The Recovery-Stress Questionnaire for Work (RESTQ-Work) [61] assesses the stress state and experienced resources in the past 7 days/nights. In this study, the short version of the RESTQ-Work (RESTQ-Work-27) with 27 items was used. The items can be assigned to a stress or resource score. The stress score consists of 10 items, and the resource score consists of 17 items. The answer scale is a 7-point scale ranging from 0 (never) to 6 (always). One example item for the stress score is "In the past 7 days/nights…I felt frustrated through my work," and one example item for the resource score is "In the past 7 days/nights…I had the chance to make suggestions at work."

Statistical Analyses

The analyses consist of two parts. First, bivariate correlations showed the relationships between all study variables. Second, a hierarchical regression analysis was used to test the hypotheses regarding the moderator effects of health-promoting leadership on the outcomes of stress and work-related resources. To test the moderating effects of health-promoting leadership, interaction terms between health-promoting leadership and all four risks in digital work variables were computed. Before computing the interaction terms, the variables were mean centered (ie, *z* standardized). For the analyses, SPSS 26.0 (IBM Corp) was used.

Results

Descriptive Statistics

Descriptive statistics (means and standard deviations) and reliabilities (Cronbach α) of the study variables are shown in Table 2. Correlations of all study variables are shown in Table 3.

Table 2. Descriptive statistics and reliabilities of the study variables.

Dimension	Score, mean (SD)	α
Distributed team work	2.44 (1.53)	.60
Mobile work	1.06 (1.19)	.62
Constant availability	2.19 (1.80)	.71
Inefficient technical support	2.43 (1.66)	.85
Health-promoting leadership	2.99 (1.59)	.93
Work-related resources	3.24 (1.04)	.92
Stress	2.01 (1.30)	.93

Table 3. Correlations between all study variables (N=1412).

Dimension	Distributed team work	Mobile work	Constant availability	Inefficient technical support	Health-promoting leadership	Work-related resources	Stress
Distributed team work							
r	1	0.10	−0.13	0.58	−0.52	−0.56	0.28
P value	—[a]	<.001	<.001	<.001	<.001	<.001	<.001
Mobile work							
r	0.10	1	0.29	0.16	−0.09	−0.10	0.22
P value	<.001	—	<.001	<.001	<.001	<.001	<.001
Constant availability							
r	−0.13	0.29	1	−0.10	0.11	0.12	0.08
P value	<.001	<.001	—	<.001	<.001	<.001	<.001
Inefficient technical support							
r	0.58	0.16	−0.10	1	−0.51	−0.50	0.26
P value	<.001	<.001	<.001	—	<.001	<.001	<.001
Health-promoting leadership							
r	−0.52	−0.09	0.11	−0.51	1	0.66	−0.38
P value	<.001	<.001	<.001	<.001	—	<.001	<.001
Work-related resources							
r	−0.56	−0.10	0.12	−0.50	0.66	1	−0.43
P value	<.001	<.001	<.001	<.001	<.001	—	<.001
Stress							
r	0.28	0.22	0.08	0.26	−0.38	−0.43	1
P value	<.001	<.001	<.001	<.001	<.001	<.001	—

[a]Not applicable.

Regression Analyses

To test our hypotheses, two step-wise regression analyses were conducted where stress and work-related resources served as the outcomes and the risk factors of digital work and health-promoting leadership served as the predictor variables. In the first step, the risk factors of digital work, including distributed team work, mobile work, constant availability, and inefficient technical support, were entered. In the consecutive second step, health-promoting leadership was entered as the moderator variable. In the third and final step, the interaction terms of the moderating variable health-promoting leadership with the four risk factors were entered. To test if multicollinearity was an issue in our data, we tested the variance inflation factor for all independent variables. All variance inflation factors were below 3 (ranging from 1.02 to 1.70). Thus, multicollinearity was not an issue in our study.

Regression Analysis With the Outcome Stress

Table 4 summarizes the regression results for the criterion stress and shows that distributed team work, mobile work, constant availability, and inefficient technical support accounted for 13% of the variance in stress. Distributed team work (β=.19, *P*<.001), mobile work (β=.15, *P*<.001), constant availability (β=.08, *P*=.003), and inefficient technical support (β=.13, *P*<.001) showed significant relationships with stress.

In the second step, health-promoting leadership accounted for an additional 6% of the variance in stress. The relationship with stress was negative (β=−.31, *P*<.001), indicating that high health-promoting leadership is associated with low stress.

In the third and final step, the interaction terms of the moderating variables were entered. The two interaction terms mobile work*health-promoting leadership (β=.11, *P*<.001) and inefficient technical support*health-promoting leadership (β=.07, *P*=.03) were significant. The results did not show a moderating effect of health-promoting leadership for the predictors distributed team work and constant availability. This step accounted for an additional 2% of the variance in stress.

Table 4. Results of hierarchical multiple regression analyses for the criterion stress (R2=20.7%).

Step and variable	Stress results[a]						
	B	SE B	β	*P* value	*t* (df)[b]	*F* (df)	*P* value
Step 1						50.880 (4,1407)	<.001
Distributed team work	0.25	0.04	.19	<.001	6.32 (1407)		
Mobile work	0.21	0.04	.15	<.001	5.74 (1407)		
Constant availability	0.11	0.04	.08	.003	2.94 (1407)		
Inefficient technical support	0.17	0.04	.13	<.001	4.17 (1407)		
Step 2						65.673 (5,1406)	<.001
Distributed team work	0.12	0.04	.09	.003	3.01 (1406)		
Mobile work	0.20	0.04	.15	<.001	5.67 (1406)		
Constant availability	0.12	0.03	.09	<.001	3.56 (1406)		
Inefficient technical support	0.04	0.04	.03	.30	1.03 (1406)		
Health-promoting leadership	−0.40	0.04	−.31	<.001	−10.45 (1406)		
Step 3						40.545 (9,1402)	<.001
Distributed team work	0.12	0.04	.09	.003	3.02 (1402)		
Mobile work	0.21	0.04	.15	<.001	5.87 (1402)		
Constant availability	0.11	0.03	.08	.001	3.23 (1402)		
Inefficient technical support	0.04	0.04	.03	.30	1.03 (1402)		
Health-promoting leadership	−0.35	0.04	−.27	<.001	−8.84 (1402)		
Distributed team work*health-promoting leadership	−0.05	0.04	−.04	.23	−1.19 (1402)		
Mobile work*health-promoting leadership	0.15	0.04	.11	<.001	4.21 (1402)		
Constant availability*health-promoting leadership	0.04	0.03	.03	.25	1.15 (1402)		
Inefficient technical support*health-promoting leadership	0.08	0.04	.07	.03	2.19 (1402)		

[a]Step 1: ΔR^2=12.64 (*P*<.001); Step 2: ΔR^2=6.30 (*P*<.001); Step 3: ΔR^2=1.72 (*P*<.001).

[b]df for *t* values were calculated with the formula N-p-1 (p=number of parameters).

Regression Analysis With the Outcome Work-Related Resources

The results for the criterion work-related resources showed that distributed team work, mobile work, constant availability, and inefficient technical support accounted for 37% of the variance in work-related resources (Table 5). Out of these four predictors, distributed team work (β=−.40, *P*<.001), constant availability (β=.05, *P*=.02), and inefficient technical support (β=−.26, *P*<.001) showed significant relationships with work-related resources. Unexpectedly, constant availability showed a low but positive relationship with work-related resources, indicating that being constantly available for work is associated with higher work-related resources. Therefore, only distributed team work and inefficient technical support were negatively related to work-related resources.

In the second step, health-promoting leadership accounted for an additional 15% of the variance in work-related resources. The relationship was positive (β=.47, *P*<.001), indicating that high health-promoting leadership is associated with higher employees' work-related resources.

In the third and final step, the interaction terms of the moderating variables were entered. The one interaction term of mobile work*health-promoting leadership (β=.04, *P*=.04) was significant. However, the results did not show a moderating effect of health-promoting leadership for the other three predictors.

Table 5. Results of hierarchical multiple regression analyses for the criterion work-related resources (R2=51.3%).

Step and variable	Work-related resources[a]						
	B	SE B	β	*P* value	*t* (df)[b]	*F* (df)	*P* value
Step 1						204.152 (4,1407)	<.001
Distributed team work	−0.42	0.03	−.40	<.001	−15.43 (1407)		
Mobile work	−0.03	0.03	−.03	.17	−1.39 (1407)		
Constant availability	0.06	0.02	.05	.02	2.39 (1407)		
Inefficient technical support	−0.27	0.03	−.26	<.001	−9.77 (1407)		
Step 2						295.090 (5,1406)	<.001
Distributed team work	−0.26	0.03	−.25	<.001	−10.39 (1406)		
Mobile work	−0.02	0.02	−.02	.31	−1.02 (1406)		
Constant availability	0.04	0.02	.03	.09	1.72 (1406)		
Inefficient technical support	−0.11	0.03	−.11	<.001	−4.56 (1406)		
Health-promoting leadership	0.49	0.02	.47	<.001	20.43 (1406)		
Step 3						166.282 (9,1402)	<.001
Distributed team work	−0.26	0.03	−.25	<.001	−10.40 (1402)		
Mobile work	−0.01	0.02	−.01	.55	−0.60 (1402)		
Constant availability	0.04	0.02	.03	.11	1.62 (1402)		
Inefficient technical support	−0.12	0.03	−.12	<.001	−4.79 (1402)		
Health-promoting leadership	0.48	0.02	.47	<.001	19.68 (1402)		
Distributed team work*health-promoting leadership	−0.03	0.02	−.03	.28	−1.09 (1402)		
Mobile work*health-promoting leadership	0.05	0.02	.04	.04	2.04 (1402)		
Constant availability*health-promoting leadership	−0.03	0.02	−.03	.10	−1.64 (1402)		
Inefficient technical support*health-promoting leadership	−0.04	0.02	−.04	.08	−1.75 (1402)		

[a]Step 1: ΔR^2=36.72 (*P*<.001); Step 2: ΔR^2=14.48 (*P*<.001); Step 3: ΔR^2=0.43 (*P*=.02).

[b]df for *t* values were calculated with the formula N-p-1 (p=number of parameters).

Simple Slope Analyses

In order to investigate the interaction effects, simple slope analyses were conducted for the significant interaction effects (Figures 2-4). The slopes indicated that employees with high health-promoting leadership experienced less stress than employees with low health-promoting leadership. However, employees with high mobile work did not seem to benefit much from health-promoting leadership, as employees with high mobile work and with high health-promoting leadership seemed to have a similar stress level as that in employees with low health-promoting leadership (Figure 2).

As for the risk factor inefficient technical support, having inefficient technical support was related to higher employee stress. Having high health-promoting leadership could buffer this negative relationship, as the stress level of these employees was lower compared to that in employees with low health-promoting leadership (Figure 3).

As for work-related resources, employees with high health-promoting leadership experienced more resources at the workplace than employees with low health-promoting leadership. In terms of low health-promoting leadership, work-related resources were the lowest in the group of employees with high mobile work (Figure 4).

Figure 2. Effect of a two-way interaction between mobile work and health-promoting leadership on stress. prom.: promoting.

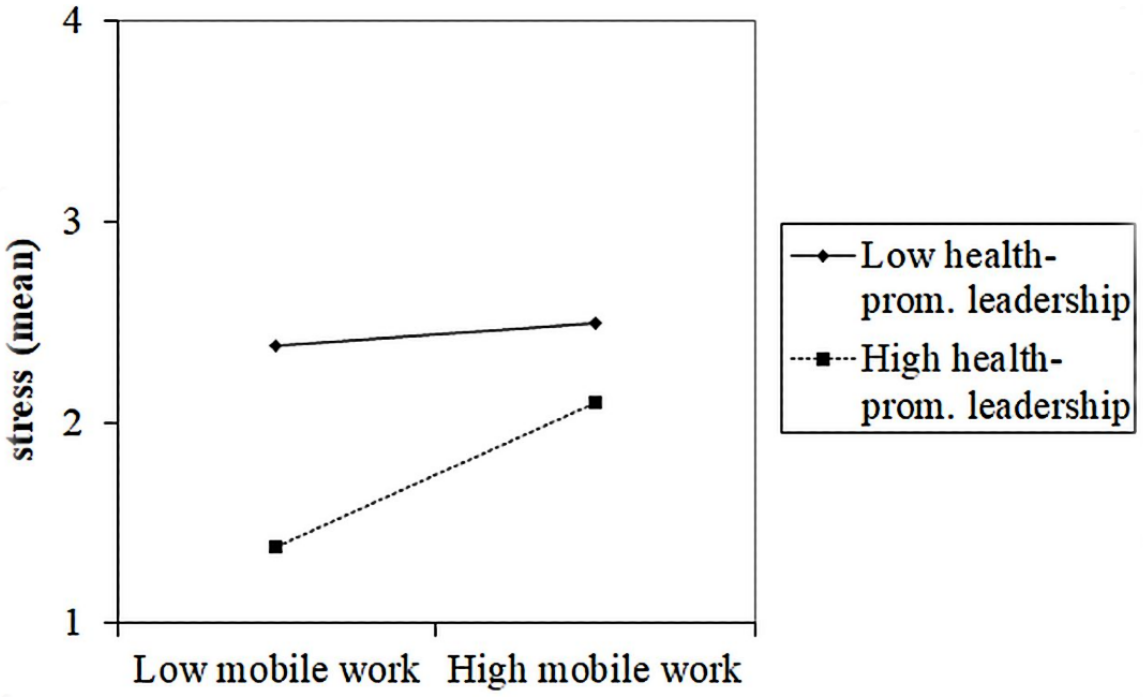


Figure 3. Effect of a two-way interaction between inefficient technical support and health-promoting leadership on stress. prom.: promoting.

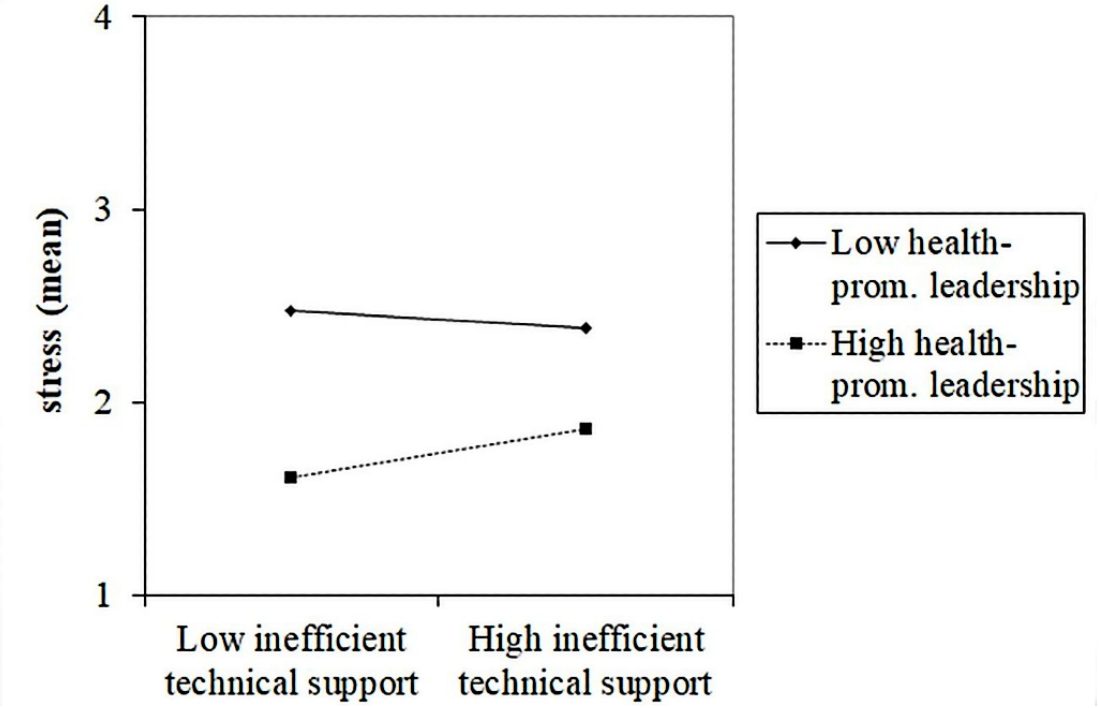
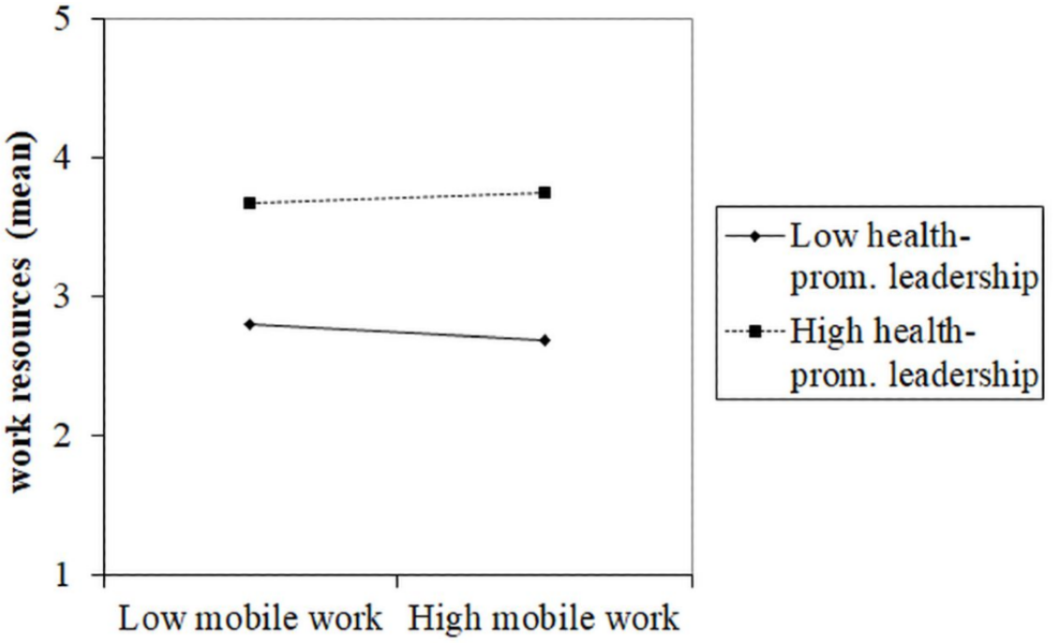


Figure 4. Effect of a two-way interaction between mobile work and health-promoting leadership on work-related resources. prom.: promoting.



Discussion

Principal Results

Risk Factors of Digital Work

This study explored the relationships between four risk factors of digital work (distributed team work, mobile work, constant availability, and inefficient technical support) and employees' stress and work-related resources. In addition, the potential role of health-promoting leadership in reducing the critical effects of digital work was investigated.

The results showed that all four risk factors of digital work (distributed team work, mobile work, constant availability, and inefficient technical support) were related to higher employee stress. In addition, distributed team work and inefficient technical support were associated with lower work-related resources.

This is in line with previous literature on digital workplaces. Distributed team work (or in other words, virtual team work) can lead to higher stress, since team collaboration and team support are difficult in teams with low face-to-face contact [20]. This might be the reason for experiencing higher stress and lower resources in this kind of teamwork, such as lower participation and decision possibilities and lower social support. Inefficient technical support, such as receiving low support in learning and using digital tools, is a critical factor as well, which has potential harmful effects on employees' well-being. Support options, such as training, peer assistance, and efficient support from the technical department if available, are factors that are relatively easy to implement in the organization and can reduce the critical effects on stress and resources.

The results for the risk factor mobile work showed a positive relationship with stress, which is consistent with previous literature. It has been shown that working in multiple locations increases mental demands, such as more interruptions or distractions, increased feeling of "timeless" continuous work, and constant changing of the rhythm of work [15,32]. However, an effect on work-related resources could not be found in this study. This means that mobile work neither increases nor decreases the work-related resources of employees. Important work-related resources, such as autonomy, decision-making, and participation opportunities, as well as social contact with colleagues, do not seem to be affected by mobile work.

We expected that being constantly available for work via telephone or email would be related to higher stress and lower work-related resources. Indeed, constant availability was related to higher employee stress, indicating that the expectation of having to be constantly available for work can lead to difficulties detaching from work, which harms the well-being of employees [35,36]. Unexpectedly, being constantly available for work showed a low but positive relationship with work-related resources. Another study conducted by ten Brummelhuis et al [4] came to a similar conclusion. In their study, constant availability via email or telephone was associated with greater work flexibility, which is perceived as a resource by employees. The simultaneous perception of increased resources and increased stress at the workplace seems implausible at first, but is actually not a contradiction. In the view of Kallus [62], increased stress and increased work-related resources can occur simultaneously. This seems to be the case with our findings in this study. Being constantly available increases stress, as employees might have difficulties detaching from work. At the same time, employees might experience higher flexibility, which is also associated with work-related resources such as higher autonomy.

However, chronic stress might tax the employees' resources to the extent that resources are damaged and lost to the point where they cannot be activated anymore [56]. Employees and organizations must therefore always pay attention to a balance between stress and resources. The relationships between constant availability and both outcomes were small though. Further

studies are needed to deepen the understanding of the possible critical and beneficial effects of constant availability.

Health-Promoting Leadership

Increasing digitalization of the workplace should support employees in their work tasks and not additionally burden them. Leaders play a key role in ensuring that work is designed in a health-promoting way [9]. In this study, we investigated if leaders engaging in health-promoting leadership could act as a buffer between the risks emerging from digital work and critical outcomes of employee well-being.

The results showed interaction effects between mobile work and health-promoting leadership, as well as between inefficient technical support and health-promoting leadership. As for the first interaction effect, the analysis revealed that the combination of low mobile work and high health-promoting leadership was related to low employee stress. This means that having a work location that usually does not change during the week and having a health-promoting leader seems to be the best condition for employee well-being. For employees with high mobile work, the beneficial effect of health-promoting leadership on stress could not be verified. It is possible that the conditions of mobile work make health-promoting leadership behavior more difficult, since the physical distance places special demands on the management and promotion of employees. Therefore, health-promoting leadership cannot buffer the critical effect on stress anymore, as leaders are far away most of the time.

Interestingly, mobile work did not increase or decrease the work-related resources of employees. A potential buffer effect of health-promoting leadership could not be verified. An explanation could be that mobile work itself is perceived as a resource by involving higher autonomy and more decision-making and participation opportunities. Physical distance can also be an obstacle for leaders to build up work-related resources [63].

Further, the results showed an interaction effect between health-promoting leadership and inefficient technical support. Experiencing high support in using digital tools and being led in a health-promoting way seems to be the best combination for employees in regard to stress. With a health-promoting leader and at the same time fewer support opportunities in learning and using digital tools, the stress of employees is high but still below that of employees with low health-promoting leadership. In other words, in the case of experiencing hindrances in learning and using digital tools, leaders can weaken the potential critical effects on stress. This is in line with previous findings, where high support from supervisors helped employees to cope with using new technology at the workplace, which demands high learning effort [42,50].

Theoretical Implications

Mental risk factors in the workplace that lead to mental stress must be carefully evaluated in each workplace according to international norms like ISO 45001 [64] and especially European laws (eg, the Framework Directive 89/391/EEC [65] and the European Framework for Psychosocial Risk Management, PRIMA-EF [66]). In this study, we were able to show that risk factors in a digitalized work environment must be considered in addition to the commonly evaluated risk factors. Currently, the so-called risk assessment focuses strongly on the following areas: the physical environment, the organizational and social environment, and the task itself (ISO 10075-1 [11]). This study presents the following four possible risk factors that could be included in addition to the aforementioned areas: distributed team work, mobile work, constant availability, and inefficient technical support. We strongly suggest including these risks in current theoretical concepts about risk assessment at the workplace.

In research regarding NWW, possible negative effects of new forms of working are already considered [7]. However, research regarding the positive aspects of NWW still outweighs research regarding the negative effects. Although the advantages of NWW are obvious, such as being flexible regarding working time and location and having higher work-family balance, negative effects are possible if the working conditions are not optimally designed. The results of this study showed that a more critical view of the effects of NWW should be included in research.

In this study, we assumed that leaders who lead in a health-promoting way act as a buffer between work-related demands and employee well-being [54]. The results showed that this buffer effect was visible for one of the four risk factors (inefficient technical support). Although stress among employees increases if they receive little support, explanation, and information when using digital tools, the increase is not as strong if leaders lead in a health-promoting way. For the other three risk factors (distributed team work, mobile work, and constant availability), the results did not indicate a buffer effect.

Employees experience the best working conditions when workplaces have low risks and when health-promoting leadership is high. In a digitalized working world with special risks, such as virtual teamwork, mobile working, and constant availability, it seems that leaders need to show leadership behavior adapted to these working conditions in order to reduce employee stress and increase work-related resources. Research is yet to define such a leadership model that is best suited for digital workplaces. Initial approaches in the field of virtual teams exist, which could serve as a base for such a leadership model [17,67,68]. Nevertheless, the goal should be a broader leadership concept that goes beyond research on virtual teams.

Practical Implications

In order to remain competitive, many companies are switching to elements of new ways of working, such as home office, mobile work, and increased use of digital media. For successful digitalization of the working process, both the company and individual employees must adapt well to the changed working conditions. Therefore, interventions to support health-promoting digitalization of the workplace have to be developed. In a workplace where digitalization is already well advanced, it is plausible to set digital interventions. For example, the whole process of workplace health promotion can be done digitally, starting with electronic feedback tools to recognize employees' health states and extending to creating and implementing eHealth tools [69]. The adoption of eHealth tools to promote physical and mental health is an effective way to support

employees [70-72]. Organizations can also benefit from eHealth tools by quickly receiving anonymized feedback about the well-being of their employees. In the event of critical feedback, the organization can act to avoid negative consequences, such as stress and burnout.

Leaders in particular must recognize the needs of their employees in a digitalized work environment even more strongly than in traditional work settings and adapt their leadership behavior accordingly. In addition, in the time of COVID-19, the support of leaders plays a much stronger role in reducing the stress for employees [73]. During the COVID-19 pandemic, many employees are working in home offices, and thus, solutions are needed for leaders on how employees can be optimally supported from a distance. The results of this study provide initial insights into the difficulties of leadership in a digitalized work environment. For example, our results indicate that when employees have high mobile work and therefore are locally distant from their leaders, leaders need more support to be able to lead in a health-promoting way. For this kind of work, certain aspects of digitalization can be an advantage, as digital tools can allow leaders to keep close contact with their employees, for example, by using video calls or chat.

Limitations

This study was a cross-sectional study with the data collected at one measurement point. To determine causality, longitudinal analyses are needed. It seems plausible that risk factors at the organization level affect the well-being of employees and not the other way around. However, it is possible that highly stressed employees perceive certain work characteristics more negatively and thus rate these characteristics as more demanding.

Same-source bias is a possible limitation of the study. As we asked employees to rate risk factors in the organization, we assessed the perceived risk factors from the view of employees. Health-promoting leadership was measured in a similar way. Although most research in the field of work-related risks and work characteristics has been conducted at the individual level (at the level of employees), a multilevel view of work characteristics (eg, bringing together the rating of teams) is a more accurate measurement of risk factors in the organization.

Since we conducted the study through an online panel organization, we did not have any personal information of the participants, such as names and email addresses. Additionally, participation did not entail any obligation or dependency. As a result, we were able to reduce fears of anonymity, and therefore, we can assume that the responses were honest. Of course, there is always the effect that people want to present themselves better than how they are in reality. We cannot completely rule out the possibility that people answered questions about their work environment more critically or less critically. However, we assume that the way the study was conducted reduced this bias.

Conclusions

The results show that all four risk factors of digital work (distributed team work, mobile work, constant availability, and inefficient technical support) are related to higher stress among employees. As for a possible buffer effect of health-promoting leadership, we found that leaders can mitigate the critical effect of inefficient technical support on stress by showing health-promoting leadership behavior. However, risk factors, such as virtual team work and mobile work, might need a different leadership behavior to reduce the health-impairing effects on employee well-being. The physical distance between leaders and employees in virtual team work and in mobile work might hamper leaders in leading in a health-promoting way. Interestingly, being constantly available for work, including during leisure time, is not as much of a risk as other factors, as employees perceive more work-related resources. More research is needed to identify the conditions under which constant availability has beneficial or impairing effects on the well-being of employees.

Acknowledgments

The authors wish to thank Agnes Diebschlag, Michaela Höfer, Cornelia Hubich-Schmon, Angelika Lepold, and Martin Weßel for their support in the organization of the study “Working World D-A-CH.” This publication has financial support from the University of Graz.

Conflicts of Interest

None declared.

References

1. Bordi L, Okkonen J, Mäkiniemi J, Heikkilä-Tammi K. Communication in the Digital Work Environment: Implications for Wellbeing at Work. NJWLS 2018 Apr 12;8(S3). [doi: 10.18291/njwls.v8is3.105275]
2. Köffer S. Designing the digital workplace of the future – what scholars recommend to practitioners. 2015 Presented at: International Conference on Information Systems (ICIS 2015); December 13-16, 2015; Fort Worth, USA.
3. Blok MM, Groenesteijn L, Schelvis R, Vink P. New Ways of Working: does flexibility in time and location of work change work behavior and affect business outcomes? Work 2012;41 Suppl 1:2605-2610. [doi: 10.3233/WOR-2012-1028-2605] [Medline: 22317114]
4. ten Brummelhuis LL, Bakker AB, Hetland J, Keulemans L. Do new ways of working foster work engagement? Psicothema 2012 Feb;24(1):113-120. [Medline: 22269373]

5. Duque L, Costa R, Dias Á, Pereira L, Santos J, António N. New Ways of Working and the Physical Environment to Improve Employee Engagement. Sustainability 2020 Aug 20;12(17):6759. [doi: 10.3390/su12176759]
6. Gerards R, de Grip A, Baudewijns C. Do new ways of working increase work engagement? PR 2018 Mar 05;47(2):517-534. [doi: 10.1108/pr-02-2017-0050]
7. Kotera Y, Correa Vione K. Psychological Impacts of the New Ways of Working (NWW): A Systematic Review. Int J Environ Res Public Health 2020 Jul 14;17(14) [FREE Full text] [doi: 10.3390/ijerph17145080] [Medline: 32674518]
8. Schaufeli WB. Engaging leadership in the job demands-resources model. Career Dev Int 2015 Sep 14;20(5):446-463. [doi: 10.1108/cdi-02-2015-0025]
9. Jiménez P, Winkler B, Bregenzer A. Developing Sustainable Workplaces with Leadership: Feedback about Organizational Working Conditions to Support Leaders in Health-Promoting Behavior. Sustainability 2017 Oct 26;9(11):1944. [doi: 10.3390/su9111944]
10. Bregenzer A, Wagner-Hartl V, Jiménez P. Who uses apps in health promotion? A target group analysis of leaders. Health Informatics J 2019 Sep;25(3):1038-1052 [FREE Full text] [doi: 10.1177/1460458217738121] [Medline: 29113522]
11. ISO 10075-1:2017 Ergonomic principles related to mental workload — Part 1: General issues and concepts, terms and definitions. URL: https://www.iso.org/standard/66900.html [accessed 2021-03-02]
12. Beehr TA, Jex SM, Stacy BA, Murray MA. Work stressors and coworker support as predictors of individual strain and job performance. J. Organiz. Behav 2000 Jun;21(4):391-405. [doi: 10.1002/(sici)1099-1379(200006)21:4<391::aid-job15>3.0.co;2-9]
13. International Labour Organization. Psychosocial Factors at Work: Recognition and Control. Geneva, Switzerland: International Labour Office; 1986.
14. Dulebohn JH, Hoch JE. Virtual teams in organizations. Human Resource Management Review 2017 Dec;27(4):569-574. [doi: 10.1016/j.hrmr.2016.12.004]
15. Vartiainen M, Hyrkkänen U. Changing requirements and mental workload factors in mobile multi‐locational work. New Technology, Work and Employment 2010;25(2):117-135. [doi: 10.1111/j.1468-005X.2010.00243.x]
16. Van Steenbergen EF, van der Ven C, Peeters MCW, Taris TW. Transitioning Towards New Ways of Working: Do Job Demands, Job Resources, Burnout, and Engagement Change? Psychol Rep 2018 Aug;121(4):736-766 [FREE Full text] [doi: 10.1177/0033294117740134] [Medline: 29298562]
17. Hoch JE, Kozlowski SWJ. Leading virtual teams: hierarchical leadership, structural supports, and shared team leadership. J Appl Psychol 2014 May;99(3):390-403. [doi: 10.1037/a0030264] [Medline: 23205494]
18. Bergiel BJ, Bergiel EB, Balsmeier PW. Nature of virtual teams: a summary of their advantages and disadvantages. Management Research News 2008 Jan 25;31(2):99-110. [doi: 10.1108/01409170810846821]
19. Hinds PJ, Bailey DE. Out of Sight, Out of Sync: Understanding Conflict in Distributed Teams. Organization Science 2003 Dec;14(6):615-632. [doi: 10.1287/orsc.14.6.615.24872]
20. Hinds PJ, Mortensen M. Understanding Conflict in Geographically Distributed Teams: The Moderating Effects of Shared Identity, Shared Context, and Spontaneous Communication. Organization Science 2005 Jun;16(3):290-307. [doi: 10.1287/orsc.1050.0122]
21. Martínez-Moreno E, Zornoza A, Orengo V, Thompson LF. The Effects of Team Self-Guided Training on Conflict Management in Virtual Teams. Group Decis Negot 2014 Nov 9;24(5):905-923. [doi: 10.1007/s10726-014-9421-7]
22. Martins LL, Gilson LL, Maynard MT. Virtual Teams: What Do We Know and Where Do We Go From Here? Journal of Management 2016 Jun 23;30(6):805-835. [doi: 10.1016/j.jm.2004.05.002]
23. Nurmi N. Coping with coping strategies: how distributed teams and their members deal with the stress of distance, time zones and culture. Stress Health 2011 Apr 13;27(2):123-143. [doi: 10.1002/smi.1327] [Medline: 27486615]
24. Friedman RA, Tidd ST, Currall SC, Tsai JC. What Goes Comes Around: The Impact of Personal Conflict Style on Work Conflict and Stress. Int Jnl of Conflict Management 2000 Jan;11(1):32-55. [doi: 10.1108/eb022834]
25. Barley SR, Meyerson DE, Grodal S. E-mail as a Source and Symbol of Stress. Organization Science 2011 Aug;22(4):887-906. [doi: 10.1287/orsc.1100.0573]
26. Hislop D, Axtell C. The neglect of spatial mobility in contemporary studies of work: the case of telework. New Tech Work Empl 2007 Mar;22(1):34-51. [doi: 10.1111/j.1468-005x.2007.00182.x]
27. Telework and ICT-based mobile work: Flexible working in the digital age. Eurofound. 2020. URL: https://www.eurofound.europa.eu/publications/report/2020/telework-and-ict-based-mobile-work-flexible-working-in-the-digital-age [accessed 2021-03-02]
28. Neis B, Lippel K. Occupational Health and Safety and the Mobile Workforce: Insights From a Canadian Research Program. New Solut 2019 Nov;29(3):297-316. [doi: 10.1177/1048291119876681] [Medline: 31608823]
29. Redman T, Snape E, Ashurst C. Location, Location, Location: Does Place of Work Really Matter? British Journal of Management 2009;20:171-181. [doi: 10.1111/j.1467-8551.2008.00640.x]
30. Hill EJ, Hawkins AJ, Ferris M, Weitzman M. Finding an Extra Day a Week: The Positive Influence of Perceived Job Flexibility on Work and Family Life Balance. Family Relations 2001 Jan;50(1):49-58. [doi: 10.1111/j.1741-3729.2001.00049.x]

31. Brunelle E. Leadership and Mobile Working: The Impact of Distance on the Superior-Subordinate Relationship and the Moderating Effects of Leadership Style. IJBSS 2013;4(11):1-14. [doi: 10.30845/ijbss]
32. Paridon H, Hupke M. Psychosocial Impact of Mobile Telework: Results from an Online Survey. EJOP 2009 Feb 28;5(1). [doi: 10.5964/ejop.v5i1.282]
33. Park J, Kim S, Lee H. Effect of work-related smartphone use after work on job burnout: Moderating effect of social support and organizational politics. Computers in Human Behavior 2020 Apr;105:106194. [doi: 10.1016/j.chb.2019.106194]
34. Büchler N, ter Hoeven CL, van Zoonen W. Understanding constant connectivity to work: How and for whom is constant connectivity related to employee well-being? Information and Organization 2020 Sep;30(3):100302. [doi: 10.1016/j.infoandorg.2020.100302]
35. Derks D, van Duin D, Tims M, Bakker AB. Smartphone use and work-home interference: The moderating role of social norms and employee work engagement. J Occup Organ Psychol 2014 Aug 29;88(1):155-177. [doi: 10.1111/joop.12083]
36. Mellner C. After-hours availability expectations, work-related smartphone use during leisure, and psychological detachment: The moderating role of boundary control. Intl J of Workplace Health Mgt 2016 Jun 13;9(2):146-164. [doi: 10.1108/ijwhm-07-2015-0050]
37. Derks D, ten Brummelhuis LL, Zecic D, Bakker AB. Switching on and off … : Does smartphone use obstruct the possibility to engage in recovery activities? European Journal of Work and Organizational Psychology 2012 Aug 28;23(1):80-90. [doi: 10.1080/1359432x.2012.711013]
38. Dettmers J. How extended work availability affects well-being: The mediating roles of psychological detachment and work-family-conflict. Work & Stress 2017 Mar 27;31(1):24-41. [doi: 10.1080/02678373.2017.1298164]
39. Demerouti E, Derks D, ten Brummelhuis L, Bakker AB. New Ways of Working: Impact on Working Conditions, Work–Family Balance, and Well-Being. In: Korunka C, Hoonakker P, editors. The Impact of ICT on Quality of Working Life. Dordrecht: Springer; 2014:123-142.
40. Stadin M, Nordin M, Broström A, Magnusson Hanson LL, Westerlund H, Fransson EI. Information and communication technology demands at work: the association with job strain, effort-reward imbalance and self-rated health in different socio-economic strata. Int Arch Occup Environ Health 2016 Oct;89(7):1049-1058 [FREE Full text] [doi: 10.1007/s00420-016-1140-8] [Medline: 27193569]
41. Rutkowski A, Saunders CS. Growing Pains with Information Overload. Computer 2010 Jun;43(6):96-95. [doi: 10.1109/mc.2010.171]
42. Salanova M, Llorens S, Cifre E. The dark side of technologies: technostress among users of information and communication technologies. Int J Psychol 2013;48(3):422-436. [doi: 10.1080/00207594.2012.680460] [Medline: 22731610]
43. Shu Q, Tu Q, Wang K. The Impact of Computer Self-Efficacy and Technology Dependence on Computer-Related Technostress: A Social Cognitive Theory Perspective. International Journal of Human-Computer Interaction 2011 Oct;27(10):923-939. [doi: 10.1080/10447318.2011.555313]
44. Weil M, Rosen L. TechnoStress: Coping with Technology @Work @Home @Play. Hoboken, NJ, USA: John Wiley & Sons Inc; 1997.
45. La Torre G, Esposito A, Sciarra I, Chiappetta M. Definition, symptoms and risk of techno-stress: a systematic review. Int Arch Occup Environ Health 2019 Jan;92(1):13-35. [doi: 10.1007/s00420-018-1352-1] [Medline: 30196317]
46. Ayyagari R, Grover V, Purvis R. Technostress: Technological Antecedents and Implications. MIS Quarterly 2011;35(4):831. [doi: 10.2307/41409963]
47. Korunka C, Kubicek B, Paškvan M, Ulferts H. Changes in work intensification and intensified learning: challenge or hindrance demands? Journal of Managerial Psych 2015 Sep 14;30(7):786-800. [doi: 10.1108/jmp-02-2013-0065]
48. Ragu-Nathan TS, Tarafdar M, Ragu-Nathan BS, Tu Q. The Consequences of Technostress for End Users in Organizations: Conceptual Development and Empirical Validation. Information Systems Research 2008 Dec;19(4):417-433. [doi: 10.1287/isre.1070.0165]
49. Yueh H, Lu M, Lin W. Employees' acceptance of mobile technology in a workplace: An empirical study using SEM and fsQCA. Journal of Business Research 2016 Jun;69(6):2318-2324. [doi: 10.1016/j.jbusres.2015.12.048]
50. Knani M, Fournier P, Biron C. Psychosocial risks, burnout and intention to quit following the introduction of new software at work. Work 2018;60(1):95-104. [doi: 10.3233/WOR-182714] [Medline: 29733035]
51. Atanasoff L, Venable MA. Technostress: Implications for Adults in the Workforce. The Career Development Quarterly 2017 Dec 07;65(4):326-338. [doi: 10.1002/cdq.12111]
52. Morrison-Smith S, Ruiz J. Challenges and barriers in virtual teams: a literature review. SN Appl. Sci 2020 May 20;2(6). [doi: 10.1007/s42452-020-2801-5]
53. Purvanova RK, Bono JE. Transformational leadership in context: Face-to-face and virtual teams. The Leadership Quarterly 2009 Jun;20(3):343-357. [doi: 10.1016/j.leaqua.2009.03.004]
54. Wegge J, Shemla M, Haslam SA. Leader Behavior as a Determinant of Health at Work: Specification and Evidence of Five Key Pathways. German Journal of Human Resource Management: Zeitschrift für Personalforschung 2014 Feb 01;28(1-2):6-23. [doi: 10.1177/239700221402800102]
55. Bakker AB, Demerouti E. The Job Demands‐Resources model: state of the art. Journal of Managerial Psych 2007 Apr 03;22(3):309-328. [doi: 10.1108/02683940710733115]

56. Hobfoll SE. Conservation of resources: A new attempt at conceptualizing stress. American Psychologist 1989;44(3):513-524. [doi: 10.1037/0003-066x.44.3.513]
57. Bregenzer A, Felfe J, Bergner S, Jiménez P. How followers' emotional stability and cultural value orientations moderate the impact of health-promoting leadership and abusive supervision on health-related resources. German Journal of Human Resource Management 2019 Feb 02;33(4):307-336. [doi: 10.1177/2397002218823300]
58. Jiménez P, Bregenzer A, Kallus KW, Fruhwirth B, Wagner-Hartl V. Enhancing Resources at the Workplace with Health-Promoting Leadership. Int J Environ Res Public Health 2017 Oct 20;14(10):1264 [FREE Full text] [doi: 10.3390/ijerph14101264] [Medline: 29053640]
59. Maslach C, Leiter MP. Early predictors of job burnout and engagement. J Appl Psychol 2008 May;93(3):498-512. [doi: 10.1037/0021-9010.93.3.498] [Medline: 18457483]
60. Jiménez P, Winkler B, Dunkl A. Creating a healthy working environment with leadership: the concept of health-promoting leadership. The International Journal of Human Resource Management 2016 Feb 19;28(17):2430-2448. [doi: 10.1080/09585192.2015.1137609]
61. Jiménez P, Dunkl A, Kallus KW. Recovery-Stress-Questionnaire for Work. In: Kallus KW, Kellmann M, editors. The Recovery-Stress Questionnaires: User Manual. Frankfurt, Germany: Pearson Assessment & Information GmbH; 2016:158-187.
62. Kallus K. Stress and Recovery: An Overview. In: Kallus KW, Kellmann M, editors. RESTQ. The Recovery-Stress Questionnaire. Frankfurt, Germany: Pearson Assessment & Information GmbH; 2016:27-48.
63. DeRosa DM, Hantula DA, Kock N, D'Arcy J. Trust and leadership in virtual teamwork: A media naturalness perspective. Hum. Resour. Manage 2004;43(2-3):219-232. [doi: 10.1002/hrm.20016]
64. ISO 45001: Occupational Health and Safety. URL: https://www.iso.org/iso-45001-occupational-health-and-safety.html [accessed 2021-03-02]
65. Council Directive of 12 June 1989 on the introduction of measures to encourage improvements in the safety and health of workers at work (89/391/EEC). EUR-Lex. URL: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:01989L0391-20081211 [accessed 2021-02-25]
66. Leka S, Jain A, Cox T, Kortum E. The development of the European framework for psychosocial risk management: PRIMA-EF. J Occup Health 2011 Apr 20;53(2):137-143 [FREE Full text] [doi: 10.1539/joh.o10010] [Medline: 21325735]
67. Liao C. Leadership in virtual teams: A multilevel perspective. Human Resource Management Review 2017 Dec;27(4):648-659. [doi: 10.1016/j.hrmr.2016.12.010]
68. Schmidt GB. Virtual Leadership: An Important Leadership Context. Ind. organ. psychol 2015 Jan 07;7(2):182-187. [doi: 10.1111/iops.12129]
69. Jimenez P, Bregenzer A. Integration of eHealth Tools in the Process of Workplace Health Promotion: Proposal for Design and Implementation. J Med Internet Res 2018 Feb 23;20(2):e65 [FREE Full text] [doi: 10.2196/jmir.8769] [Medline: 29475828]
70. Balk-Møller NC, Poulsen SK, Larsen TM. Effect of a Nine-Month Web- and App-Based Workplace Intervention to Promote Healthy Lifestyle and Weight Loss for Employees in the Social Welfare and Health Care Sector: A Randomized Controlled Trial. J Med Internet Res 2017 Apr 10;19(4):e108 [FREE Full text] [doi: 10.2196/jmir.6196] [Medline: 28396303]
71. Carolan S, Harris PR, Cavanagh K. Improving Employee Well-Being and Effectiveness: Systematic Review and Meta-Analysis of Web-Based Psychological Interventions Delivered in the Workplace. J Med Internet Res 2017 Jul 26;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]
72. De Cocker K, De Bourdeaudhuij I, Cardon G, Vandelanotte C. The Effectiveness of a Web-Based Computer-Tailored Intervention on Workplace Sitting: A Randomized Controlled Trial. J Med Internet Res 2016 May 31;18(5):e96 [FREE Full text] [doi: 10.2196/jmir.5266] [Medline: 27245789]
73. Evanoff BA, Strickland JR, Dale AM, Hayibor L, Page E, Duncan JG, et al. Work-Related and Personal Factors Associated With Mental Well-Being During the COVID-19 Response: Survey of Health Care and Other Workers. J Med Internet Res 2020 Aug 25;22(8):e21366 [FREE Full text] [doi: 10.2196/21366] [Medline: 32763891]

Abbreviations

ICT: information and communication technology
NWW: new ways of working

Edited by G Eysenbach; submitted 09.10.20; peer-reviewed by F Solsona, D Tchir, B Nievas Soriano; comments to author 02.11.20; revised version received 13.11.20; accepted 18.01.21; published 12.03.21.

Please cite as:
Bregenzer A, Jimenez P
Risk Factors and Leadership in a Digitalized Working World and Their Effects on Employees' Stress and Resources: Web-Based Questionnaire Study
J Med Internet Res 2021;23(3):e24906
URL: https://www.jmir.org/2021/3/e24906
doi:10.2196/24906
PMID:33709933

Original Paper

An Interactive Web-Based Sexual Health Literacy Program for Safe Sex Practice for Female Chinese University Students: Multicenter Randomized Controlled Trial

Janet Yuen-Ha Wong[1*], PhD; Wen Zhang[1*], MNurs; Yongda Wu[1], PhD; Edmond Pui Hang Choi[1], PhD; Herman Hay Ming Lo[2], PhD; Wendy Wong[3], PhD; Jasmine Hin Man Chio[4], PhD; Hau Lin Cherry Tam[5], PhD; Fei Wan Ngai[6], PhD; Marie Tarrant[7], PhD; Man Ping Wang[1], PhD; Hextan Yuen-Sheung Ngan[8], PhD; Daniel Yee-Tak Fong[1], PhD

[1]School of Nursing, University of Hong Kong, Hong Kong SAR, Hong Kong

[2]Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hong Kong SAR, Hong Kong

[3]School of Chinese Medicine, The Chinese University of Hong Kong, Hong Kong SAR, Hong Kong

[4]Department of Counselling and Psychology, Hong Kong Shue Yan University, Hong Kong SAR, Hong Kong

[5]Department of Social and Behavioural Science, City University of Hong Kong, Hong Kong SAR, Hong Kong

[6]School of Nursing, The Hong Kong Polytechnic University, Hong Kong SAR, Hong Kong

[7]School of Nursing, The University of British Columbia, Vancouver, BC, Canada

[8]Department of Obstetrics and Gynaecology, University of Hong Kong, Hong Kong SAR, Hong Kong

[*]these authors contributed equally

Corresponding Author:
Daniel Yee-Tak Fong, PhD
School of Nursing
University of Hong Kong
Pokfulam
Hong Kong SAR
Hong Kong
Phone: 852 39176645
Email: dytfong@hku.hk

Abstract

Background: Sexual health concerns among young adults worldwide help to motivate preventative practices against sexually transmitted infections. To foster better sexual health, sexual health literacy must be enhanced. Little research has been conducted on the impact of gender power dynamics on sexual health, such as sexual coercion, even though the prevalence of sexual coercion remains high in China.

Objective: This study describes the development and systematic evaluation of a web-based sexual health literacy intervention called “Smart Girlfriend” for female Chinese university students.

Methods: A multicenter randomized controlled trial was conducted with 781 female university students at 5 universities with dormitories in Hong Kong. Inclusion criteria were used to select unmarried, female, Chinese university students who were ≥18 years old and had not received a sexual health intervention in the past 12 months. Participants were randomly assigned to 2 groups: one group received an interactive web-based sexual health literacy intervention and the other group received a single webpage of online information about condom use. The intervention content was based on the Health Belief Model and the Continuum of Conflict and Control theory. The primary outcome was self-reported consistency of condom use with every partner at 3-month and 6-month follow-up assessments, analyzed using zero/one inflated beta (ZOIB) regression. The secondary outcome was an appraisal of the knowledge, attitudes, norms, and self-efficacy of condom use using the 25-item Multidimensional Condom Attitudes Scale (MCAS). The intention to treat was applied in analyses.

Results: Of 1503 individuals that were screened, 781 (52%) were randomized into 2 groups. The retention rates at the 3-month and 6-month follow-ups were 92% and 91%, respectively. Most participants were born locally (536/746, 72%), and 18% (134/746) self-reported as a sexual minority. ZOIB results regarding the consistency of condom use were not significant [model 1: odds ratio (OR) 2.25 with a 95% credible interval (CrI) of 0.84-6.36; model 2: OR 8.03 (95% CrI 0.22-330.31); model 3: OR 1.21 (95% CrI 0.78-1.86)]. Consistency in the intervention group was 5% higher (95% CI −1.90 to 11.63) than the control group at

the 3-month follow-up, and 1% higher (95% CI −5.81 to 8·02) at the 6-month follow-up. MCAS scores at the 3-month follow-up were significantly higher in the intervention group (mean 122.51, SD 15.97) than the control group (mean 119.86, SD 15.85; *P*=.02).

Conclusions: An interactive web-based sexual health literacy program did not significantly increase the consistency of condom use compared to a single webpage of condom use information; however, it did temporarily improve knowledge, attitudes, norms, and self-efficacy regarding condom use. Future revisions of this intervention should be personalized and delivered with a proactive approach.

Trial Registration: ClinicalTrials.gov NCT03695679; https://clinicaltrials.gov/ct2/show/NCT03695679

(J Med Internet Res 2021;23(3):e22564) doi:10.2196/22564

KEYWORDS

sexual health; eHealth; women's health; sex education; health literacy

Introduction

According to the World Health Organization (WHO), sexual health is a state of physical, mental, and social well-being in relation to sexuality [1]. Sexually healthy individuals have an absence of sexual or reproductive disease and a positive approach to managing respectful sexual relationships free of coercion and violence, thereby exhibiting safe sexual behaviors. Sexual health helps prevent sexually transmitted infections (STIs) and the human papillomavirus (HPV), the second leading cause of cancer deaths among women globally [2], which the WHO proposed as a key global health sector strategy for 2016-2021 [3].

Previous studies have indicated the important role of health literacy in promoting sexual health [4-6]. Health literacy is a form of empowerment to enhance individuals' capacity to access, understand, communicate, and process health information and services to make appropriate health decisions [7]. Sexual health literacy is the ability to understand preventive sexual health information to make informed choices, increase safe sex practices (eg, promoting condom use, limiting the number of sexual partners, avoiding causal sex, and enhancing sexual communication and negotiation skills with regard to sex refusal, condom use, and a partner's STI history), and reduce STI risk [4]. Low sexual health literacy is related to poor sexual health decision-making among university students, including engaging in risky sexual behaviors (such as inconsistent condom use) and delays or difficulties in seeking care [5]. Young adults might generally have good health and thus may not fully understand the importance of sexual health assessment in the absence of obvious symptoms, particularly after engaging in risky sexual behaviors [6]. Due to increasing trends of premarital sex and unsafe sexual behaviors [8], technologically advanced dating apps [9,10], and engagement in compensatory dating and casual sex [11], the risks of STIs and cervical cancer in female university students is high. Therefore, it is paramount that female university students have adequate sexual health literacy to facilitate safe sex practices, which yields better sexual health.

In this study, we operationalize sexual health literacy as individuals' capacity to understand risk information about unsafe sex practices and communicate with sexual partners to make optimal decisions and maintain sexual health. Previous sexual health literacy interventions focused on specific groups of people, such as a 50-minute face-to-face interactive class for HIV-positive people [12] and a 10-hour face-to-face intervention for jailed females to prevent cervical cancer [13]. However, few interventions targeted young females in the general population to prevent common STIs. Moreover, regarding the intervention content, it has been increasingly recognized that interventions should go beyond biophysical content, such as human development and contraception skills, by also including sociocultural content, such as respectful relationships and sexual coercion [14]. Research on effective interventions addressing sexual coercion and safe sex is limited, even though freedom from sexual coercion is a key conceptual component of good sexual health [15].

The internet was the most commonly accessed source for sexual health information in young adults [16]. A web-based sexual health intervention has some potential advantages over a traditional face-to-face intervention, including the capacity to reach a larger number of people in the population with a relatively low cost and facilitating communication with full privacy and confidentiality [16]. Moreover, a web-based intervention's interactive and anonymous delivery is more acceptable and effective [17]; a Cochrane meta-analysis evaluating 15 randomized controlled trials (RCTs) on safe-sex practice interventions found the interactive computer-based interventions to be more effective at improving knowledge about sexual health [18]. In this study, we have developed an interactive web-based sexual health literacy intervention to promote safe sex practices.

The sexual health status of young women in China is generally poor [19], and the level of sexual health knowledge was poorer than what was reported for young women in Western countries [20]. Young women are also vulnerable to risky sexual behaviors [21]. A national survey found that 1.6% of female Chinese university students reported having multiple sexual partners [22], and another study found that only 17.2% of sexually active female Chinese university students consistently used condoms [23]. The prevalence of sexual coercion of female university students in Hong Kong was reported to be approximately 13% in 2015, showing no reduction since 2008 [24,25]. Sexual health interventions in Hong Kong have lagged far behind that of many other places [26]. Moreover, sexual health interventions in Hong Kong often focus primarily on physiological knowledge and the dissemination of STI and STI-prevention information; little attention is paid to the effects of gender-power dynamics on

sexual health, such as sexual coercion, respectful relationships, and sexual communication and negotiation [26]. However, a growing body of literature shows that the more sexual communication and negotiation that occurs before sex, the more likely a condom will be used during sex [27]. Sexual coercion is highly related to unsafe sex practices [28]. Taken together, these facts emphasize a need to revisit, develop, and evaluate comprehensive sexual health literacy interventions in the Chinese context.

In this study, we describe the development and systematic evaluation of a sexual health literacy program called "Smart Girlfriend" using a multicenter RCT. This program is an interactive web-based intervention that aims to disseminate knowledge about STIs and condom use, communication and negotiation about condom use, and sexual coercion in daily life to enhance safe sex practices among female university students in Hong Kong.

Methods

Trial Design

The design was a multicenter, single-blind, parallel-group RCT, registered with ClinicalTrials.gov (NCT03695679). Ethical approval was obtained from the institutional review board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (UW-17029) and other universities. We followed the CONSORT (Consolidated Standards of Reporting Trials) guidelines.

Sample Size

Our sample size calculation was based on a primary comparison of behavioral change in the consistency with condom use. A previous study reported that the mean percentage of condom use was 68% (SD 39%) in individuals who underwent a computer-based intervention, compared with 24% (SD 35%) among those in a control condition, corresponding to a moderate to large Cohen effect size of 0.6 [29]. To conservatively detect a small Cohen effect size of 0.3 with 80% power and a maximum 5% false positive error rate using a 2-sided, 2-sample *t* test, a total of 352 (ie, 2 groups of 176) female university students were required. Assuming a 30% attrition rate based on a previous study using a web-based intervention [30], we calculated a total necessary sample size of 500 participants.

Participants

In total, 1503 female university students across various disciplines and years of study were screened from 5 universities with dormitory or residential halls in Hong Kong. Of these, 781 students were recruited. The eligibility criteria were (1) female university students aged 18 years or above who (2) are able to read Chinese or understand Cantonese, (3) are unmarried, (4) have been with intimate partners (ie, current and former dating partners or partners in a relationship) in the past 12 months, and (5) have not received any sexual health information (including formal face-to-face or online education or training courses related to contraceptives and sexually transmitted diseases, from a university, hospital, clinic, or nongovernmental organization) in the past 12 months. The exclusion criteria were (1) an unwillingness to complete the questionnaires at 3 time-points, (2) being pregnant or postnatal, and (3) having a psychiatric illness.

Recruitment, Randomization, Masking, and Allocation Concealment

Students were reached via leaflets, campus booths, bulk emails, and posters. All interested participants received an invitation email to log into the Smart Girlfriend webpage (Multimedia Appendix 1-3). Online enrolment was used to screen students for eligibility. For eligible participants, written informed consent was obtained online, followed by a baseline questionnaire. After completing the questionnaire, the recruited participants were randomized to either the intervention group or the control group, according to a list prepared by blocked randomization (with blocks of 4), with a 1:1 randomization ratio. The block size and order of allocation were kept securely in the randomizer to avoid selection bias. The online platform conducted masking and allocation concealment according to the participants' enrolment sequence. Participants were automatically guided to the webpage associated with their allocation and were not aware of the group allocation in advance. Participants who completed all of the web-based questionnaires were given vouchers with a value of HKD $300 (USD $38.50). The privacy of all participants was ensured. Data collected from all questionnaires were stored in a protected university database.

Previous Development Efforts of Smart Girlfriend

Findings from our previous local multisite survey (n=1076) revealed that Chinese university students saw the relationship between unsafe sex practices and sexual coercion as problematic [25]. The results showed that 12% of female university students who were engaged in a dating relationship or had dating partners in the past year experienced sexual coercion within that year [25]. Therefore, we conducted an intervention to help enhance the awareness of sexual coercion among university students. Sexual issues are typically taboo subjects in Chinese culture; therefore, to minimize embarrassing situations for participants, we implemented the Dating Compassion, Assessment, reFerral, and Education (Dating CAFE) Ambassador Programme to help Chinese university students with dating violence. The intervention's development was based on the theory of planned behaviors and was conducted via 3 face-to-face workshops (totaling 7.5 hours) with 81 university students. Compared to students in the control group, we found that the students trained to be ambassadors had significantly enhanced behavioral intentions and control to help peers who were experiencing dating violence, decreased acceptance of dating violence, and increased subjective norms for helping others [31]. Moreover, we learned that discussions about dating issues were attractive and acceptable among university students. In addition, the face-to-face intervention was labor-intensive, and some students could not be enrolled due to timetable clashes. Thus, a web-based intervention could be a more cost-effective and practical approach for reaching as many eligible young people as possible.

The Smart Girlfriend Intervention

The development of Smart Girlfriend was based on the Dating CAFE initiative. Smart Girlfriend is designed to be a sexual

health literacy intervention empowering female university students with enhanced knowledge, attitudes, norms, and self-efficacy around managing sexual health, particularly condom use for safe sex practice. Based on the Health Belief Model (Figure 1) [32], we delivered Smart Girlfriend in 3 phases. In the first phase, participants were able to check their perceived susceptibility and perceived severity for individual STIs. The Harvard Cancer Risk Index [33] was used to ask 8 questions to estimate individual STI and cervical cancer risks. Respondents obtained personalized results from their answers, including factors that may increase or reduce their risk of getting STIs and cervical cancer. In the second phase, participants were given knowledge-based information about STIs and cervical cancer in text format, including relevant statistics, development, possible symptoms, and prevention methods. Some scale-based questions were asked to help the students think about the positive and negative features of condom use. Participants were able to offer personal feedback and were provided with an opportunity to reflect on which benefits or barriers mattered most to them (Figure 2).

Figure 1. Conceptual framework of the Smart Girlfriend program based on the Health Belief Model. STIs: sexually transmitted infections.

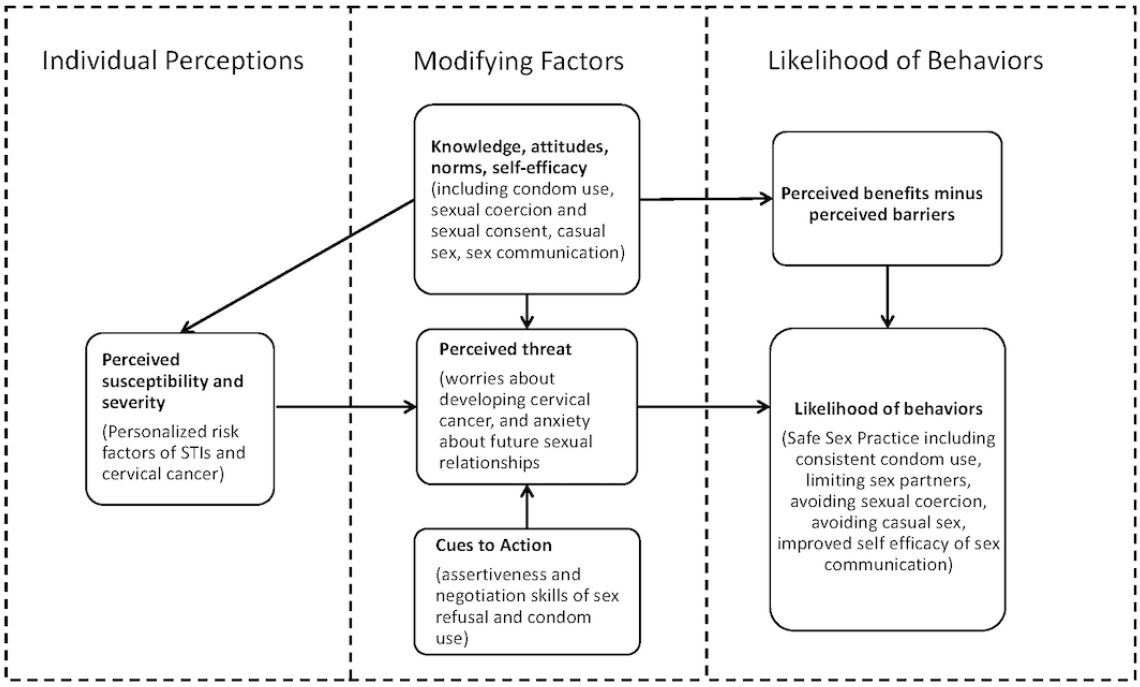


Figure 2. An example of a scale-based question asked by Smart Girlfriend to help students think about the positive and negative features of condom use and provide an opportunity for feedback and reflection on what matters most to them. STIs: sexually transmitted infections.

Use condom				Not use condom		
I want to do everything I can to prevent STIs.				I am the first and only sexual partner for him/her. The chance of getting STIs is very low.		
☐	☐	☐	☐	☐	☐	☐
More important			Equally important			More important

In the second phase, participants were also prompted to take action regarding condom use. Therefore, condom use procedures and tips, as well as web links for local STI testing, cervical screening programs, and HPV vaccination programs, were provided. In addition, participants' self-efficacy in condom use was enhanced by providing information about assertiveness in sexual consent and sexual communication to avoid sexual coercion and casual sex. Three 5-minute videos were created with narrative stories about STIs and HPV infection based on different scenarios relevant to common situations for university students, about the handling of sexual consent in dormitories, engaging in sexual communication at home, and talking to a friend about worries regarding sex without a condom after a Christmas party (Multimedia Appendix 2). The Continuum of Conflict and Control (CCC) theory [34] was used to guide the stories, emphasizing that sexual coercion can occur without

physical violence and with minimal fear to strengthen participants' knowledge about sexual coercion, sexual consent pertaining to condom use, and sex refusal. Participants were able to assess their own values and receive feedback on their choices. Textbox 1 shows the perceived benefits of and barriers to condom use featured in the interactive intervention.

Textbox 1. Perceived benefits of and barriers to condom use featured in the interactive intervention.

Perceived benefits of condom use:

- Consistent use of latex condoms before having sex lowers the risk of cervical cancer.
- Consistent and correct condom use is necessary every time one has sex, from the start of the act to the finish, to effectively lower the risk of sexually transmitted infections (STIs).
- A condom acts as a barrier against human papillomavirus (HPV) and other STIs.
- Some types of HPV can cause cells in the cervix to become cancerous; a condom acts as a barrier against HPV and other STIs.

Perceived barriers to condom use:

- Not knowing the risk of STIs [Link to knowledge about STIs and cervical cancer, including its statistics, development, and possible symptoms]
- Not knowing how to use a condom [Link to condom use procedures and tips]
- Low comfort levels with initiating conversations about condom use
- Feelings of frustration when being rejected due to issues around condom use [Link to video 3 about feelings and provide information to connect condom use and health risks of STIs and cervical cancer]
- Not knowing what to do after being rejected due to issues around condom use
- Sexual coercion [Link to definition and statistics of sexual coercion and examples]

In the last phase, a page designed to summarize the participants' individual factors for facilitating decision-making about consistent condom use in future sexual activities was presented. Participants were asked to rate their level of self-efficacy in terms of knowledge, skills, clarity of information, and perceived support and advice on a scale of 1-5. If the participant's level of self-efficacy was low (1-3), they were directed to relevant information via hyperlinks.

In the control group, participants received minimal intervention, with only a single webpage of online information about procedures and tips for condom use. The site for the control group had a similar graphic design to the one used for the intervention group, but no self-assessment material or online quiz questions were presented.

An inquiry system was created to handle questions from the participants. This system was designed to help participants in both the intervention and control groups if they needed any support or wanted to seek further clarification.

The time spent engaging with the online information was approximately 30 minutes for the intervention group and 10 minutes for the control group.

Outcomes and Follow-up

All data were collected from the webpage. Participants were sent a reminder email and SMS text message for completing the online questionnaires at 3 time-points: baseline (T1); 3-months postintervention (T2); and 6-months postintervention (T3). The primary outcomes were the consistency of condom use with every partner, in accordance with the recommended guidelines of a systematic review of 56 studies [35], which used the percentage of male condom protected sex with every partner during the past 3 months.

Secondary outcomes were (1) knowledge, attitudes, norms, and self-efficacy of condom use, as appraised by the 25-item Multidimensional Condom Attitudes Scale (MCAS) [36]; (2) knowledge, attitudes, norms, and self-efficacy of sexual coercion and sexual consent, as measured by the 39-item Sexual Consent Scale–Revised (SCS-R) [37]; and (3) self-efficacy in sexual communication, estimated by the 20-item Sexual Communication Self-efficacy Scale (SCSES) [38]. The MCAS items were answered using a 7-point Likert scale, and total scores ranged from 7 to 175; a previous study has shown acceptable validity and reliability in the Chinese population [39]. The Cronbach alpha in this study was .84. The SCS-R contained 3 attitudinal subscales (positive attitude toward establishing consent, lack of perceived behavioral control, and sexual consent norms) and 2 behavioral subscales (indirect consent behaviors, and awareness and discussion). The SCS-R items were answered using a 7-point Likert scale, and the Cronbach alpha in this study was .67. The SCSES items were answered using a 4-point Likert scale, and total scores ranged from 20 to 80. With the exception of one of the SCS-R subscales (lack of perceived behavioral control), high scores on the scales indicated a high level of measured outcomes. The Cronbach alpha in this study was .94.

Other outcomes included participant inquiries, participant satisfaction, and participation in the intervention. Participants' inquiries were collected to understand their further needs. Satisfaction with the intervention was evaluated by recording the overall satisfaction of the intervention on a scale of 0-10. In addition, participants were asked which part of the intervention was most memorable. The higher the score, the higher the overall satisfaction with the intervention. The total recorded views of each website page (recorded via Google Analytics) and the incidence of searching for more information were measured.

Requested demographic information included sexual orientation, birthplace, dating relationship status, and history of childhood sexual coercion. Information regarding individual risk of STIs and cervical cancer (Multimedia Appendix 3) was collected at baseline, including age, age at first sexual intercourse experience, number of sexual partners during one's lifetime, frequency of condom use, history of being diagnosed with an STI, smoking status, history of giving birth, and history of having a Pap smear test. The experience of sexual coercion was measured using a 7-item subscale of the Revised Conflict Tactic Scale [40], which indicated whether it had happened and how often the behavior had occurred in the past year.

Statistical Analysis

The primary outcome, consistency of condom use, was analyzed using a zero/one inflated beta (ZOIB) regression model because the raw data of the consistency of condom use were not normally distributed, exhibiting excessive zeros and ones (ie, consistency of condom use=100%). ZOIB is based on a piecewise distribution, which accounts for the probability mass at 0 or 1 and the probability density within (0,1). Bayesian-based results were obtained using Stan (Stan Development Team; 4 chains, 4000 iterations, 1500 warm-ups) [41]. Odds ratios (ORs) and credible interval (CrI) values were calculated for Bayesian-based analysis.

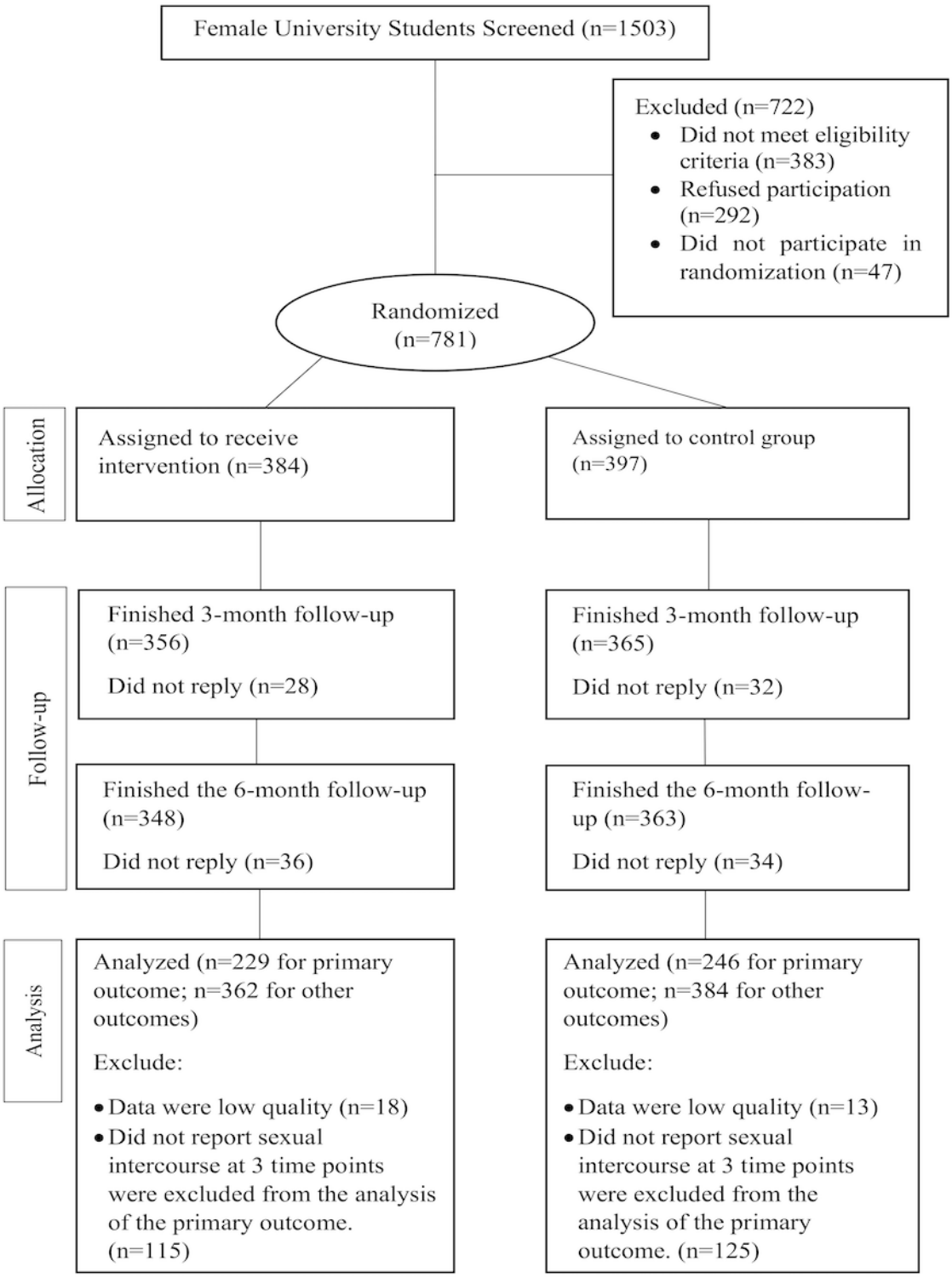
We adopted the intention-to-treat principle, and all study subjects were included in the analysis. Missing values at the 3- and 6-month follow-ups were replaced by the last observed value. Missing values at baseline were replaced by values from the 3- or 6-month follow-up. If there was no value obtained at any of the 3 time-points, the participant was excluded from the analysis. For other outcomes, the *t* test was applied for continuous data and the chi-square test was applied for categorical data. All P values were 2-sided, and P<.05 was considered statistically significant. R software (version 3.6.1; R Core Team) with “tidyr” and “brms” packages was used to analyze the data. Questions were downloaded from the inquiry system. Content analysis was performed to categorize the collected responses.

Results

CONSORT Flowchart

Figure 3 shows the flow of participants at baseline, grouping randomization, 3-month follow-up, and 6-month follow-up. Of 1503 screened students, 722 were excluded and 781 students were enrolled (enrolment rate: 52%), randomized into the intervention group (384/781, 49%) and the control group (397/781, 51%). Of the 781 included participants, the dropout rate was 8% (60/781) at the 3-month follow-up (intervention: 28/384, 7%; control: 32/397, 8%) and 9% (70/781) at the 6-month follow-up (intervention: 36/384, 9%; control: 34/397, 9%). Finally, 229 participants in the intervention group and 246 participants in the control group were included in the analysis of the primary outcome. A total of 362 participants in the intervention group and 384 participants in the control group were analyzed for other outcomes. Details of the quality check criteria used are provided in Multimedia Appendix 4.

Figure 3. CONSORT flow diagram.



Baseline Characteristics of Participants

The mean age of participants was 21.5 (SD 2.6) years, ranging from 18 to 32 years. Most participants were born locally (536/746, 72%), and 18% (134/746) self-reported as sexual minorities (Table 1). Approximately 10% (78/746) of participants reported child sexual abuse experiences, and 22% (164/746) reported a history of sexual coercion.

Table 1. Characteristics of participants at baseline (n=746).

Characteristics	Total participants (n=746)	Intervention group (n=362)	Control group (n=384)
Age in years, mean (SD)	21.5 (2.6)	21.5 (2.7)	21.5 (2.6)
Birthplace, n (%)			
Hong Kong	536 (72)	252 (70)	284 (74)
Other	210 (28)	110 (30)	100 (26)
Smoking status, n (%)			
Do not smoke	674 (90)	329 (91)	345 (90)
Quit smoking	65 (9)	29 (8)	36 (9)
Engages in smoking	7 (1)	4 (1)	3 (1)
Alcohol consumption status, n (%)			
Do not consume alcohol	309 (42)	150 (41)	159 (42)
Quit consuming alcohol	129 (17)	58 (16)	71 (19)
Consumes alcohol	308 (41)	154 (43)	154 (40)
Sexual orientation, n (%)			
Heterosexual[a]	612 (82)	294 (81)	318 (83)
Sexual minority[b]	134 (18)	68 (20)	66 (17)
History of child sexual abuse, n (%)			
Occurred	668 (90)	323 (89)	345 (90)
Did not occur	78 (10)	39 (11)	39 (10)
History of sexual coercion, n (%)			
Occurred	164 (22)	77 (21)	87 (23)
Did not occur	580 (78)	285 (79)	295 (77)
History of sexually transmitted infections, n (%)			
No	741 (99)	358 (99)	383 (99)
Yes	5 (1)	4 (1)	1 (1)
History of Pap smear test, n (%)			
Has not had a Pap smear test	702 (94)	339 (94)	363 (95)
Has had a Pap smear test	44 (6)	23 (6)	21 (6)
Relationship status, n (%)			
Dating/in a relationship	597 (80)	280 (77)	317 (83)
Broken up/ cohabiting	149 (20)	82 (23)	67 (17)
Sexual experience, n (%)			
Has not had a sexual experience with a partner	293 (39)	141 (39)	152 (40)
Has had a sexual experience with a partner	453 (61)	221 (61)	232 (60)
Age of first sexual intercourse experience, in years (n=463[c]), mean (SD)	19.6 (2.5)	19.6 (2.5)	19.7 (2.6)
Number of sexual partners (n=463[c]), mean (SD)	2.1 (2.8)	2.2 (3.3)	2.0 (2.1)

[a]For the purposes of this study, *heterosexual* was defined as an individual who is exclusively attracted to the opposite sex.

[b]For the purposes of this study, due to small respective sample sizes, the category of *sexual minority* included all respondents who reported that their sexual orientation is "mostly heterosexual," "bisexual," "mostly homosexual," "completely homosexual," and "not sure."

[c]This item only includes respondents who reported sexual experience at baseline.

Outcomes

Consistency of Condom Use

ZOIB modeling revealed nonsignificant differences between the intervention and control groups in all models regarding the consistency of condom use (Table 2). The intervention group showed a nonsignificant trend toward being more likely to report 0% or 100% consistency of condom use compared to the control group (OR 2.25, 95% CrI 0.84-6.36), and of those, a nonsignificant trend toward being more likely to report 100% condom use consistency compared to the control group (OR 8.03, 95% CrI 0.22-330.31). At baseline, the consistency of condom use was 78% in the intervention group and 72% in the control group (95% CI −0.97 to 13.04). At the 3-month follow-up, the intervention group exhibited 5% higher consistency (intervention: 80% vs control: 75%, 95% CI −1.90 to 11.63) than the control group, and exhibited 1% higher consistency (intervention: 77% vs control: 76%, 95% CI −5.1 to 8.02) at the 6-month follow-up.

Table 2. Impact of the Smart Girlfriend program on the consistency of condom use at different time points.

Model, group, and time-point	Estimate (SE[a])	OR[b]	95% CrI[c]
Model 1: Report 0% or 100% condom use consistency in all participants			
Group	0.81 (0.52)	2.25	0.84-6.36
Time3M[d]	0.06 (0.33)	1.06	0.56-2.02
Time6M[e]	0.17 (0.33)	1.18	0.62-2.25
Group* Time3M[f]	0.00 (0.48)	1.00	0.39-2.58
Group* Time6M[g]	–0.04 (0.49)	0.96	0.37-2.48
Model 2: Report 100% condom use consistency in those reporting 0% or 100%			
Group	2.08 (1.87)	8.03	0.22-330.31
Time3M	0.85 (0.73)	2.35	0.58-10.55
Time6M	1.18 (0.74)	3.26	0.80-14.91
Group* Time3M	–0.66 (1.12)	0.52	0.06-4.57
Group* Time6M	–2.66 (1.15)	0.07	0.006-0.62
Model 3: Report a condom use consistency between 0% and 100% in all participants			
Group	0.19 (0.22)	1.21	0.78-1.86
Time3M	–0.00 (0.15)	1.00	0.74-1.34
Time6M	–0.04 (0.16)	0.96	0.70-1.31
Group* Time3M	0.07 (0.24)	1.07	0.67-1.71
Group* Time6M	0.01 (0.25)	1.01	0.63-1.67

[a]SE: standard error.

[b]OR: odds ratio.

[c]CrI: credible interval for a Bayesian-based analysis.

[d]Time3M: the 3-month follow-up.

[e]Time6M: the 6-month follow-up.

[f]Group*Time3M: the interaction effect between the group and the 3-month follow-up.

[g]Group*Time6M: the interaction effect between the group and the 3-month follow-up.

Secondary Outcomes

The MCAS scores of the intervention group significantly increased at the 3-month follow-up compared to the control group (intervention: 122.51, SD 15.97, vs control: 119.86, SD 15.85; *P*=.02; Table 3). No significant difference was found in other secondary outcomes.

Table 3. The impact of the Smart Girlfriend program on secondary outcomes at different time points.

Outcome and time-point	Intervention group (n=362), mean (SD)	Control group (n=384), mean (SD)	*t* test (df[a]=744)	*P* value
Multidimensional Condom Attitudes Scale (MCAS)				
Baseline	120.24 (SD 15.39)	118.60 (SD 16.13)	1.42	0.16
3-month follow-up	122.51 (SD 15.97)	119.86 (SD 15.85)	2.27	0.02
6-month follow-up	123.55 (SD 15.76)	121.80 (SD 16.75)	1.47	0.14
Sexual Consent Scale–Revised (SCS-R) subscale 1[b]				
Baseline	32.76 (SD 12.16)	32.89 (SD 11.93)	–0.15	0.88
3-month follow-up	31.31 (SD 11.77)	32.04 (SD 11.43)	–0.87	0.39
6-month follow-up	30.91 (SD 11.40)	30.68 (SD 11.54)	0.28	0.78
SCS-R subscale 2[c]				
Baseline	59.43 (SD 10.44)	59.27 (SD 10.43)	0.22	0.83
3-month follow-up	59.18 (SD 9.83)	58.22 (SD 9.93)	1.34	0.18
6-month follow-up	58.82 (SD 9.45)	59.31 (SD 10.75)	–0.67	0.51
SCS-R subscale 3[d]				
Baseline	28.63 (SD 5.80)	28.03 (SD 5.90)	1.42	0.16
3-month follow-up	27.80 (SD 5.64)	28.01 (SD 5.86)	–0.48	0.63
6-month follow-up	27.92 (SD 5.50)	27.96 (SD 5.77)	–0.09	0.93
SCS-R subscale 4[e]				
Baseline	34.64 (SD 6.25)	34.99 (SD 6.13)	–0.77	0.44
3-month follow-up	34.21 (SD 6.21)	34.37 (SD 6.27)	–0.34	0.73
6-month follow-up	33.88 (SD 6.27)	34.71 (SD 6.27)	–1.81	0.07
SCS-R subscale 5[f]				
Baseline	16.04 (SD 5.31)	16.29 (SD 5.28)	0.65	0.51
3-month follow-up	16.61 (SD 5.32)	16.22 (SD 5.08)	1.02	0.31
6-month follow-up	17.26 (SD 5.02)	16.88 (SD 5.25)	1.02	0.31
Sexual Communication Self-efficacy Scale (SCSE)				
Baseline	62.01 (SD 10.94)	60.54 (SD 11.26)	1.81	0.07
3-month follow-up	63.41 (SD 11.09)	62.08 (SD 10.73)	1.66	0.10
6-month follow-up	63.56 (SD 10.76)	62.80 (SD 10.77)	0.95	0.34

[a]df: degrees of freedom.

[b]SCS-R subscale 1: (lack of) perceived behavioral control.

[c]SCS-R subscale 2: positive attitude toward establishing consent.

[d]SCS-R subscale 3: indirect behavioral approach to consent.

[e]SCS-R subscale 4: sexual consent norms.

[f]SCS-R subscale 5: awareness and discussion.

Inquiry System

Of the 781 participants, 10% (81/781) sent inquiries via the webpage. Of these, 27% (22/81) of inquiries asked about condoms and other contraceptive methods, 28% (23/81) asked about STIs, 14% (11/81) asked about sexual behaviors, and 31% (25/81) asked about suggestions for intervention, technical issues, or other issues (Table 4). A significant difference was found for condoms and other contraceptive methods (χ^2=9.78, P=.002) and STIs (χ^2=6.71, *P*=.01) between the 2 groups. Compared with the control group, participants in the intervention group were more likely to send inquiries about STIs (intervention: 18 vs control: 5), whereas intervention group participants were less likely to send inquiries about condoms and other contraceptive methods (intervention: 6 vs control: 16). In the category of STIs, 3 participants in the intervention group asked whether there is a risk of STIs and how to prevent STIs among women who have sex with women.

Table 4. Inquiries sent by participants in the intervention and control groups.

Inquired issue	Total (n=81), n (%)	Intervention group (n=45), n (%)	Control group (n=36), n (%)	χ^2	*P* value
Condom and other contraceptive methods	22 (27)	6 (13)	16 (44)	9.78[a]	.002
Sexually transmitted infections	23 (28)	18 (40)	5 (14)	6.71[a]	.01
Sex life	11 (14)	8 (18)	3 (8)	1.52	.22
Technical issue	6 (7)	2 (5)	4 (12)	1.30	.25
Suggestions for the website	8 (10)	5 (11)	3 (8)	0.17	.68
Others	11 (14)	6 (13)	5 (14)	0.01	.94

[a]*P*<.01.

Satisfaction and Participation of Participants

A slight but statistically nonsignificant difference in satisfaction (t_{677}=0.15; *P*=.89) was found between the intervention (n=353) and control (n=326) groups, with average satisfaction scores of 6.19 (SD 2.75) and 6.22 (SD 2.78), respectively. Among participants who reported the type of content they remembered most (Table 5), almost half of the participants in the intervention group (48/118, 41%) and more than half of the participants in the control group (58/107, 54%) reported that content related to condoms was most memorable. In addition, more intervention group participants reported that content about sexual consent was the most memorable (32/118, 27%) than control group participants (21/107, 20%). There was a significant difference between the 2 groups for seeking out more information about safe sex practice at the 3-month follow-up (intervention: 120/334, 36% vs control: 92/353, 26%; χ^2=7.83, *P*=.005) but not at the 6-month follow-up (intervention: 105/326, 32% vs control: 97/353, 28%; χ^2=1.81, *P*=.18). Data from Google Analytics revealed that the most visited pages of the intervention were the inquiry page and the page with information about risk factors for cervical cancer.

Table 5. Content that participants reported to be the most memorable.

Most memorable content	Intervention group (n=118), n (%)	Control group (n=107), n (%)
Cervical cancer	4 (3.3)	0 (0)
Sexually transmitted infections	8 (6.8)	1 (0.9)
Sexual coercion	3 (2.5)	1 (0.9)
Condoms	48 (40.7)	58 (54.2)
Sexual consent	32 (27.2)	21 (19.7)
Other	23 (19.5)	26 (24.3)

Discussion

Principal Findings

This multicenter RCT in Hong Kong is the first study to examine the effects of an interactive web-based sexual health literacy program to promote safe sex practice among female Chinese university students. The consistency of condom use increased over time in both groups. The intervention group exhibited 5% higher consistency at the 3-month follow-up and 1% higher consistency at the 6-month follow-up compared to the control group; however, these differences were not statistically significant. Thus, the results did not reveal an effect from the intervention for increasing the consistency of condom use.

One of the possible explanations for the lack of positive effect may be due to the passive exposure of the intervention to our target audiences. In our intervention, the main components were delivered passively, including the knowledge about STIs, the narrative stories about STIs, and the introduction about sexual coercion. A meta-analysis for HIV-prevention interventions also revealed that participants in active interventions (such as client-tailored counseling and other activities to improve behavioral skills) reported a greater behavioral change improvement compared to passive interventions (such as messages to share procondom information and norms and to verbally model skills); and likewise, passive interventions in the intervention groups did not differ significantly from the control groups [42]. These findings show the positive effect of active interventions and suggest that increasing active components would be necessary for future research in promoting safe sex practice.

Another potential reason for the lack of positive effect could be associated with the nonpersonalized nature of the web-based intervention content. According to the inquiries we received from the participants, some participants asked if women who had sex with women were at risk for STIs, and some intervention content was not applicable to them. Previous studies found that personalized web-based interventions were more effective than nonpersonalized interventions, and the difference between the nonpersonalized intervention group and the blank control group

was nonsignificant [43]. Thus, behavioral change may be improved if we tailor the intervention content to the participant's characteristics, such as sexual orientation.

Additionally, the imbalanced ceiling effect for 2 groups might also contribute to the insignificant intervention effect. The proportion of participants reporting 100% condom use at baseline in the intervention group was significantly higher than the control group (68% vs 59%, respectively; χ^2=3.90, *P*=.048. Owing to the ceiling effect, the potential capacity of improvement in terms of condom use in the intervention group would be less than the control group, and the effect of intervention might be underestimated [44].

Increased MCAS scores at the 3-month follow-up indicate that the Smart Girlfriend program improved participants' knowledge, attitudes, norms, and self-efficacy of condom use temporarily; however, there was no significant long-term effect on behavior observed at the 6-month follow-up. These results indicate that the intervention could temporarily improve knowledge, attitudes, norms, and self-efficacy of condom use in young adults but cannot ultimately change their behaviors. These findings were in line with a previous study [45]. Starosta et al [46] found that a web-based intervention improved attitudes toward condoms but could not change condom use at the 3-month follow-up. Thus, it would seem that improvement occurs, but it would decrease over time if there is no further exposure in the intervention [47].

Strengths and Limitations

The strengths of the study include the detailed development of the intervention based on theory and previous experience. The low dropout rate in this study indicates that university students welcomed the intervention and that it served the needs of the target population. With the participation of 5 universities in Hong Kong, the study included a diversity of student population characteristics, sexual orientations, and sexual coercion experiences. ZOIB was employed in this study, given the nonnormal distribution data with excessive zeros and ones. In promoting sexual health literacy, self-reported condom use is characterized by a substantial number of zeros or ones and discrete nonzero counts [48]. Excessive numbers of zeros or ones in this dependent variable make it difficult to fit the data using traditional methods, including ordinary least squares models and mixed linear models, which may yield biased estimates of the results. The beta distribution is considered a versatile function that fits a broad range of probability distribution shapes. This study may serve as a promising reference for future studies with similar problems. A high participation rate (91%) was reported in this study, indicating a high acceptability of digital interventions among university students and suggesting that a large number of university students can be reached at a very low cost.

This study has implications for future studies. Temporary improvement was observed in the knowledge, attitudes, norms, and self-efficacy of condom use, but not observed postintervention in behavioral change; this suggests that future research should more thoroughly consider novel methods that maintain the intervention effect and increase the behavioral change effect of web-based sexual health literacy. First, an active approach for intervention delivery may lead to the greatest increase in condom use. For example, virtual reality's effectiveness for behavioral skills training and practical exercises has been demonstrated in other research fields [49]. Second, the intervention content should be more personalized, especially for our target participants, before the intervention. More specific and advanced information (eg, whether nonpenetrative sex is a risk factor for contracting STIs), more customized content (eg, for women who have sex with women), and personal online counseling through the website might be needed. Third, more attention should be paid to the organization of the external links to other websites. We provided external hyperlinks to other websites to give participants more opportunities to explore other related information. In hindsight, providing external hyperlinks might not have been ideal since they may have distracted the participants from the intervention pathway and led them away from (and discouraged the return to) our website [50]. Recent research tested a web-based basic version alongside a version with added links to external resources, and it was found that the latter version was not effective [51]. The external links could be listed in the last (or a separate) page in future revised interventions.

Our study has several limitations. First, despite our relatively large sample of participants, compared with many other sexual health interventions, our study might lack power concerning the primary outcome for the consistency of condom use. The number of sexually active participants during the 6-month follow-up was relatively lower than expected. We included those who only reported sexual activities in one of 3 questionnaires and used a conservative approach that considered sexually inactive participants as exhibiting no change. This conservative approach might bias our results. Second, there is a potential influence of the Hawthorne effect. A systematic review found that the Hawthorne effect could influence participants' behavior in RCTs [52]. In our study, self-reported data were collected, and participants knew that the program was designed to increase their consistency of condom use. Third, we did not collect information about the duration of participation. However, due to the digital intervention design, all participants received information about the whole intervention before completing the questionnaires. Finally, our results were vulnerable to differential error because self-reported data were collected. The intervention group may have misreported the consistency of condom use to a greater extent than the control group at baseline (intervention: 78% vs control: 72%), thereby biasing the estimated intervention effect. There is no alternative to self-reporting in web-based trials because collecting biological data via web-based systems is unfeasible. Differential error could not be avoided; however, our trial was conducted under conditions that maximize accurate reporting, using an anonymous web-based questionnaire that assured participants of confidentiality with an absence of investigators.

Conclusion

Among university students in Hong Kong, an interactive web-based sexual health literacy program resulted in a small but statistically nonsignificant increase in the consistency of condom use, as well as a significant and temporary increase in

knowledge, attitudes, norms, and self-efficacy of condom use, but not in sexual coercion, sexual consent, or sexual communication, when compared with participants receiving only one webpage of condom use information. The number of participants at enrolment and the high participation rate highlight the need for sexual health literacy programs for young adults. Moreover, a future revision of this intervention should be personalized and delivered with an active approach.

Acknowledgments

We gratefully thank The Food and Health Bureau for granting us the opportunity to conduct this project with the provision of the Health and Medical Research Fund (reference number 14150971). The study's sponsors did not play a role in the study design, data collection, data analysis, data interpretation, or report writing. We also thank all the students who participated in the study, the IT assistants from the School of Nursing at The University of Hong Kong for maintaining the website, and all university administrators for their heartfelt support and assistance in making the data collection successful.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Smart Girlfriend: an interactive web-based sexual health literacy program for safe sex practice in female Chinese university students.
[PNG File , 351 KB - jmir_v23i3e22564_app1.png]

Multimedia Appendix 2
Smart Girlfriend Screenshot: 5-minute videos with narrative stories about sexually transmitted infections (STIs) and human papillomavirus (HPV) infection.
[PNG File , 207 KB - jmir_v23i3e22564_app2.png]

Multimedia Appendix 3
Smart Girlfriend Screenshot: information collection of individual risk of sexually transmitted infections (STIs) and cervical cancer.
[PNG File , 167 KB - jmir_v23i3e22564_app3.png]

Multimedia Appendix 4
Details of the data quality check.
[DOCX File , 14 KB - jmir_v23i3e22564_app4.docx]

Multimedia Appendix 5
CONSORT-EHEALTH (V 1.6.1) checklist.
[PDF File (Adobe PDF File), 382 KB - jmir_v23i3e22564_app5.pdf]

References

1. World Health Organization. Defining sexual health: Report of a technical consultation on sexual health, 28-31 January. Geneva: World Health Organization; 2006. URL: https://www.who.int/reproductivehealth/topics/gender_rights/defining_sexual_health.pdf [accessed 2020-11-25]
2. Ferlay J, Shin H, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010 Dec 15;127(12):2893-2917 [FREE Full text] [doi: 10.1002/ijc.25516] [Medline: 21351269]
3. World Health Organization. Human papillomavirus (HPV) and cervical cancer. Sexually transmitted infections (STIs) 2016. URL: http://www.who.int/mediacentre/factsheets/fs380/en/ [accessed 2020-11-25]
4. Vamos CA, Thompson EL, Logan RG, Griner SB, Perrin KM, Merrell LK, et al. Exploring college students' sexual and reproductive health literacy. J Am Coll Health 2020 Jan;68(1):79-88. [doi: 10.1080/07448481.2018.1515757] [Medline: 30388946]
5. Fortenberry JD, McFarlane MM, Hennessy M, Bull SS, Grimley DM, St Lawrence J, et al. Relation of health literacy to gonorrhoea related care. Sex Transm Infect 2001 Jun;77(3):206-211 [FREE Full text] [doi: 10.1136/sti.77.3.206] [Medline: 11402232]
6. Massey PM, Prelip M, Calimlim BM, Quiter ES, Glik DC. Contextualizing an expanded definition of health literacy among adolescents in the health care setting. Health Educ Res 2012 Dec;27(6):961-974 [FREE Full text] [doi: 10.1093/her/cys054] [Medline: 22623619]

7. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
8. Yip PSF, Zhang H, Lam T, Lam KF, Lee AM, Chan J, et al. Sex knowledge, attitudes, and high-risk sexual behaviors among unmarried youth in Hong Kong. BMC Public Health 2013 Jul 29;13:691 [FREE Full text] [doi: 10.1186/1471-2458-13-691] [Medline: 23895326]
9. Yeo TED, Ng YL. Sexual risk behaviors among apps-using young men who have sex with men in Hong Kong. AIDS Care 2016;28(3):314-318. [doi: 10.1080/09540121.2015.1093597] [Medline: 26444390]
10. Choi EPH, Wong JYH, Fong DYT. The use of social networking applications of smartphone and associated sexual risks in lesbian, gay, bisexual, and transgender populations: a systematic review. AIDS Care 2016 Jul 25;29(2):145-155. [doi: 10.1080/09540121.2016.1211606]
11. Lee TY, Shek DTL. Compensated dating in Hong Kong: prevalence, psychosocial correlates, and relationships with other risky behaviors. J Pediatr Adolesc Gynecol 2013 Jun;26(3 Suppl):S42-S48. [doi: 10.1016/j.jpag.2013.03.014] [Medline: 23683826]
12. Robinson C, Graham J. Perceived Internet health literacy of HIV-positive people through the provision of a computer and Internet health education intervention. Health Info Libr J 2010 Dec;27(4):295-303 [FREE Full text] [doi: 10.1111/j.1471-1842.2010.00898.x] [Medline: 21050372]
13. Ramaswamy M, Lee J, Wickliffe J, Allison M, Emerson A, Kelly PJ. Impact of a brief intervention on cervical health literacy: A waitlist control study with jailed women. Prev Med Rep 2017 Jun;6:314-321 [FREE Full text] [doi: 10.1016/j.pmedr.2017.04.003] [Medline: 28435785]
14. Gonçalves TR, Faria ER, Carvalho FTD, Piccinini CA, Shoveller JA. Behavioral interventions to promote condom use among women living with HIV: a systematic review update. Cad Saude Publica 2017 Jan 23;33(1):e00202515 [FREE Full text] [doi: 10.1590/0102-311X00202515] [Medline: 28125130]
15. Edwards WM, Coleman E. Defining sexual health: a descriptive overview. Arch Sex Behav 2004 Jun;33(3):189-195. [doi: 10.1023/B:ASEB.0000026619.95734.d5] [Medline: 15129038]
16. Minichiello V, Rahman S, Dune T, Scott J, Dowsett G. E-health: potential benefits and challenges in providing and accessing sexual health services. BMC Public Health 2013;13:790 [FREE Full text] [doi: 10.1186/1471-2458-13-790] [Medline: 23987137]
17. Baxter L, Egbert N, Ho E. Everyday health communication experiences of college students. J Am Coll Health 2008;56(4):427-436. [doi: 10.3200/JACH.56.44.427-436] [Medline: 18316288]
18. Bailey JV, Murray E, Rait G, Mercer CH, Morris RW, Peacock R, et al. Interactive computer-based interventions for sexual health promotion. Cochrane Database Syst Rev 2010(9):CD006483. [doi: 10.1002/14651858.CD006483.pub2] [Medline: 20824850]
19. Yu J. Teenage sexual attitudes and behaviour in China: a literature review. Health Soc Care Community 2012 Nov;20(6):561-582. [doi: 10.1111/j.1365-2524.2011.01054.x] [Medline: 22404303]
20. Chen M, Liao Y, Liu J, Fang W, Hong N, Ye X, et al. Comparison of Sexual Knowledge, Attitude, and Behavior between Female Chinese College Students from Urban Areas and Rural Areas: A Hidden Challenge for HIV/AIDS Control in China. Biomed Res Int 2016;2016:8175921 [FREE Full text] [doi: 10.1155/2016/8175921] [Medline: 28101513]
21. Wang B, Davidson P. Sex, lies, and videos in rural China: a qualitative study of women's sexual debut and risky sexual behavior. J Sex Res 2006 Aug;43(3):227-235. [doi: 10.1080/00224490609552321] [Medline: 17599245]
22. Sun X, Liu X, Shi Y, Wang Y, Wang P, Chang C. Determinants of risky sexual behavior and condom use among college students in China. AIDS Care 2013;25(6):775-783. [doi: 10.1080/09540121.2012.748875] [Medline: 23252705]
23. Ma Q, Ono-Kihara M, Cong L, Pan X, Xu G, Zamani S, et al. Behavioral and psychosocial predictors of condom use among university students in Eastern China. AIDS Care 2009 Feb;21(2):249-259. [doi: 10.1080/09540120801982921] [Medline: 19229696]
24. Chan KL, Straus MA. Prevalence and Correlates of Physical Assault on Dating Partners. TOSSCIJ 2008 Oct 30;1(1):5-14. [doi: 10.2174/1874945300801010005]
25. Wong JYH, Choi EPH, Lo HHM, Wong W, Chio JHM, Choi AWM, et al. Intimate Partner Sexual Violence and Mental Health Indicators Among Chinese Emerging Adults. J Interpers Violence 2019 Sep 03:886260519872985. [doi: 10.1177/0886260519872985] [Medline: 31478438]
26. Leung H, Lin L. Adolescent Sexual Risk Behavior in Hong Kong: Prevalence, Protective Factors, and Sex Education Programs. J Adolesc Health 2019 Jun;64(6S):S52-S58. [doi: 10.1016/j.jadohealth.2018.12.007] [Medline: 31122550]
27. Xiao Z, Li X, Lin D, Jiang S, Liu Y, Li S. Sexual communication, safer sex self-efficacy, and condom use among young Chinese migrants in Beijing, China. AIDS Educ Prev 2013 Dec;25(6):480-494. [doi: 10.1521/aeap.2013.25.6.480] [Medline: 24245595]
28. Agardh A, Odberg-Pettersson K, Ostergren P. Experience of sexual coercion and risky sexual behavior among Ugandan university students. BMC Public Health 2011 Jul 04;11:527 [FREE Full text] [doi: 10.1186/1471-2458-11-527] [Medline: 21726433]
29. Kiene SM, Barta WD. A brief individualized computer-delivered sexual risk reduction intervention increases HIV/AIDS preventive behavior. J Adolesc Health 2006 Sep;39(3):404-410. [doi: 10.1016/j.jadohealth.2005.12.029] [Medline: 16919803]

30. Warmerdam L, van Straten A, Twisk J, Riper H, Cuijpers P. Internet-based treatment for adults with depressive symptoms: randomized controlled trial. J Med Internet Res 2008;10(4):e44 [FREE Full text] [doi: 10.2196/jmir.1094] [Medline: 19033149]
31. Wong JY, Tang NR, Yau JH, Choi AW, Fong DY. Dating CAFE Ambassador Programme: Chinese College Students to Help Peers in Dating Violence. Health Educ Behav 2019 Dec;46(6):981-990. [doi: 10.1177/1090198119867736] [Medline: 31431078]
32. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [Medline: 6392204]
33. Colditz GA, Atwood KA, Emmons K, Monson RR, Willett WC, Trichopoulos D, et al. Harvard report on cancer prevention volume 4: Harvard Cancer Risk Index. Risk Index Working Group, Harvard Center for Cancer Prevention. Cancer Causes Control 2000 Jul;11(6):477-488. [doi: 10.1023/a:1008984432272] [Medline: 10880030]
34. Carlson R, Dayle Jones K. Continuum of Conflict and Control: A Conceptualization of Intimate Partner Violence Typologies. The Family Journal 2010 May 21;18(3):248-254. [doi: 10.1177/1066480710371795]
35. Noar SM, Cole C, Carlyle K. Condom use measurement in 56 studies of sexual risk behavior: review and recommendations. Arch Sex Behav 2006 Jun;35(3):327-345. [doi: 10.1007/s10508-006-9028-4] [Medline: 16799837]
36. Helweg-Larsen M, Collins BE. The UCLA Multidimensional Condom Attitudes Scale: documenting the complex determinants of condom use in college students. Health Psychol 1994 May;13(3):224-237. [Medline: 8055858]
37. Humphreys TP, Brousseau MM. The sexual consent scale-revised: development, reliability, and preliminary validity. J Sex Res 2010 Sep;47(5):420-428. [doi: 10.1080/00224490903151358] [Medline: 19685367]
38. Quinn-Nilas C, Milhausen RR, Breuer R, Bailey J, Pavlou M, DiClemente RJ, et al. Validation of the Sexual Communication Self-Efficacy Scale. Health Educ Behav 2016 Apr;43(2):165-171. [doi: 10.1177/1090198115598986] [Medline: 26286296]
39. Choi EPH, Fong DYT, Wong JYH. The use of the Multidimensional Condom Attitude Scale in Chinese young adults. Health Qual Life Outcomes 2020 Oct 08;18(1):331 [FREE Full text] [doi: 10.1186/s12955-020-01577-9] [Medline: 33032622]
40. Straus MA, Hamby SL, Boney-McCoy S, Sugarman DB. The Revised Conflict Tactics Scales (CTS2). Journal of Family Issues 2016 Jun 30;17(3):283-316. [doi: 10.1177/019251396017003001]
41. Bürkner P. Advanced Bayesian Multilevel Modeling with the R Package brms. The R Journal 2018;10(1):395. [doi: 10.32614/rj-2018-017]
42. Albarracín D, Gillette JC, Earl AN, Glasman LR, Durantini MR, Ho M. A test of major assumptions about behavior change: a comprehensive look at the effects of passive and active HIV-prevention interventions since the beginning of the epidemic. Psychol Bull 2005 Nov;131(6):856-897 [FREE Full text] [doi: 10.1037/0033-2909.131.6.856] [Medline: 16351327]
43. Mevissen FEF, Ruiter RAC, Meertens RM, Zimbile F, Schaalma HP. Justify your love: testing an online STI-risk communication intervention designed to promote condom use and STI-testing. Psychol Health 2011 Feb;26(2):205-221. [doi: 10.1080/08870446.2011.531575] [Medline: 21318930]
44. McCambridge J, Kypri K, Elbourne D. In randomization we trust? There are overlooked problems in experimenting with people in behavioral intervention trials. J Clin Epidemiol 2014 Mar;67(3):247-253 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.09.004] [Medline: 24314401]
45. Sun WH, Wong CKH, Wong WCW. A Peer-Led, Social Media-Delivered, Safer Sex Intervention for Chinese College Students: Randomized Controlled Trial. J Med Internet Res 2017 Dec 09;19(8):e284 [FREE Full text] [doi: 10.2196/jmir.7403] [Medline: 28793980]
46. Starosta AJ, Cranston E, Earleywine M. Safer sex in a digital world: A Web-based motivational enhancement intervention to increase condom use among college women. J Am Coll Health 2016;64(3):184-193. [doi: 10.1080/07448481.2015.1107835] [Medline: 26700487]
47. Sales JM, DiClemente RJ, Davis TP, Sullivan S. Exploring why young African American women do not change condom-use behavior following participation in an STI/HIV prevention intervention. Health Educ Res 2012 Dec;27(6):1091-1101 [FREE Full text] [doi: 10.1093/her/cys059] [Medline: 22641793]
48. Gupta R, Szczesniak RD, Macaluso M. Modeling repeated count measures with excess zeros in an epidemiological study. Ann Epidemiol 2015 Aug;25(8):583-589. [doi: 10.1016/j.annepidem.2015.03.011] [Medline: 25887702]
49. Hadley W, Houck C, Brown LK, Spitalnick JS, Ferrer M, Barker D. Moving Beyond Role-Play: Evaluating the Use of Virtual Reality to Teach Emotion Regulation for the Prevention of Adolescent Risk Behavior Within a Randomized Pilot Trial. J Pediatr Psychol 2019 May 01;44(4):425-435 [FREE Full text] [doi: 10.1093/jpepsy/jsy092] [Medline: 30551157]
50. Huizingh EKRE, Hoekstra JC. Why do consumers like websites? J Target Meas Anal Mark 2003 May 1;11(4):350-361. [doi: 10.1057/palgrave.jt.5740090]
51. Peels DA, van Stralen MM, Bolman C, Golsteijn RHJ, de Vries H, Mudde AN, et al. The differentiated effectiveness of a printed versus a Web-based tailored physical activity intervention among adults aged over 50. Health Educ Res 2014 Oct;29(5):870-882. [doi: 10.1093/her/cyu039] [Medline: 24980023]
52. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]

Abbreviations

CrI: credible interval
Dating CAFE: Dating Compassion, Assessment, reFerral, and Education
HPV: human papillomavirus
MCAS: Multidimensional Condom Attitudes Scale
OR: odds ratio
RCT: randomized controlled trial
SCSES: SCSES: Sexual Communication Self-efficacy Scale
SCS-R: Sexual Consent Scale-Revised
STI: sexually transmitted infection
WHO: World Health Organization
ZOIB: zero/one inflated beta

Edited by G Eysenbach; submitted 22.07.20; peer-reviewed by CY Lin, V Minichiello, PKH Mo; comments to author 03.10.20; revised version received 27.11.20; accepted 12.12.20; published 12.03.21.

Please cite as:
Wong JYH, Zhang W, Wu Y, Choi EPH, Lo HHM, Wong W, Chio JHM, Tam HLC, Ngai FW, Tarrant M, Wang MP, Ngan HYS, Fong DYT
An Interactive Web-Based Sexual Health Literacy Program for Safe Sex Practice for Female Chinese University Students: Multicenter Randomized Controlled Trial
J Med Internet Res 2021;23(3):e22564
URL: https://www.jmir.org/2021/3/e22564
doi:10.2196/22564
PMID:33709941

Review

Web-Based Dietary and Physical Activity Intervention Programs for Patients With Hypertension: Scoping Review

Fatimah Najihah Baderol Allam[1], BSc; Mohd Ramadan Ab Hamid[1], BSc, MSc; Siti Sabariah Buhari[1], BSc, MSc, PhD; Harrinni Md Noor[2], BEd, PGD, MEd, PhD

[1]Centre for Dietetics Studies, Faculty of Health Sciences, Universiti Teknologi MARA, Puncak Alam, Malaysia

[2]Faculty of Education, Universiti Teknologi MARA, Puncak Alam, Malaysia

Corresponding Author:
Mohd Ramadan Ab Hamid, BSc, MSc
Centre for Dietetics Studies
Faculty of Health Sciences
Universiti Teknologi MARA
Puncak Alam
Malaysia
Phone: 60 332584424
Email: ramadan7230@uitm.edu.my

Abstract

Background: Hypertension is the root cause of many chronic diseases. Lifestyle changes (ie, dietary alterations and physical activity) are seen to be an important step in the prevention and treatment of hypertension. Educating people through web-based interventional programs could offer an effective solution and help these patients with hypertension in the existing health care scenario.

Objective: In this study, the researchers conducted a scoping literature review of the web-based dietary changes and physical activity–related intervention programs designed for the patients with hypertension and identified the methodologies, effectiveness, protocols, and theories, which could affect and improve existing clinical activities.

Methods: This review followed the scoping review methodology to identify and process the peer-reviewed studies published between 2010 and 2020. The literature searches were conducted on the following electronic databases: PubMed, Web of Science, MEDLINE (Medical Literature Analysis and Retrieval System Online), ScienceDirect, Scopus, and Google Scholar. By using relevant search terms, studies were included if they offered information related to the web-based intervention tools, specifically dietary and physical activity intervention for patients with hypertension. Studies written or translated in English language and published within the date range (January 2010 to March 2020) were included.

Results: Overall, 1441 articles were initially identified. The reviewers included 35 articles after removing duplicates and screening titles. Only 21 articles were assessed for full review, and 15 were kept for analysis. The researchers selected 15 web-based intervention articles published on the topic of hypertension from 7 countries. A few of these 15 web-based tools (4, 27%) included more than 3 functions and provided a lot of important information (such as appointments, health records, or viewable care). Several tools were standalone tools (11, 73%), while most of the tools supported communication intervention–related lifestyle or behavioral changes (13, 87%) and medication adherence (6, 40%). It was found that physicians (9, 60%), allied health professionals (5, 33%), and nurses (5, 33%) were the health care providers who generally used these tools for communicating with their patients. More than half of the above tools (10, 67%) were assessed by different researchers in randomized controlled trials, while 5 tools (33%) were investigated in nonrandomized studies.

Conclusions: We identified many web-based intervention programs for patients with hypertension from the literature databases. The findings indicate that numerous benefits can be derived after using a web-based dietary and physical activity intervention program for hypertension focusing on lifestyle changes. However, developers need to consider the preferences of the patients with regard to the information or the design features while developing or modifying web-based educational websites. These tools could be used for designing a patient-tailored website intervention program that is based on diet and physical activities for patients with hypertension.

(J Med Internet Res 2021;23(3):e22465) doi:10.2196/22465

KEYWORDS

hypertension; blood pressure; education; website-based; dietary intake; physical activity

Introduction

Background

Hypertension, or high blood pressure, is the root cause of several chronic diseases and affected more than 31.1% of the adults (1.39 billion) around the world in 2010 [1]. One study stated that in the last 20 years, the high-income countries showed a slight decrease in the prevalence of high blood pressure, whereas the low- and middle-income countries (LMICs) showed a significantly higher prevalence of the disease. These differences in the trend have increased the burden of hypertension and other related cardiovascular issues on the health care systems of these LMICs [2].

The prevalence of this health issue has increased around the globe due to many lifestyle changes like an unhealthy diet (which is low in potassium and high in sodium) and a lack of physical activity. Furthermore, many factors such as occupation, abdominal obesity, body mass index, education levels, tobacco and alcohol usage, and the socioeconomic status of the people also contributed to hypertension [2,3]. An increased prevalence of high blood pressure is also responsible for other major diseases like diabetes, cardiovascular problems, and chronic kidney disease [4-6]. Fortunately, blood pressure can be controlled by improving lifestyle changes through the emergence of educational programs [7].

The Effect of Web-Based Educational Interventions in Lifestyle Modification

In a previous study, Ali et al stated that educational programs are an effective technique that can be used by patients with hypertension to increase their knowledge about lifestyle changes, controlling their harmful lifestyle habits, and enhancing their self-management [7]. They observed that after using these tools for 3 months, there was a significant decrease in the number of excessive-salt consumers, physically inactive people, and those who consumed an inadequate amount of fish and fiber. This indicated that the tools helped in improving the lifestyle scores of the patients. By engaging with educational intervention as one approach, this can be effective in promoting physical activity and improving weight management, nutrition, and mental health in patients with hypertension [8].

Since there is an increase in the global prevalence of hypertension, people need to undergo an urgent transformation. In the past few years, there has been an increase in many digital health-related interventions, such as mobile apps and web-based and software-based interventions. These tools allow the patients to access all health-related information and encourage them to follow lifestyle and behavioral changes [9,10]. Web-based educational interventions are suitable to be used as a medium to promote the exchange of patient education information through online mechanisms such as websites [11]. This medium of interventions allows for flexibility in the type and amount of information people can access on a routine basis [12]. This type of web-based intervention also can encourage patients to manage their high blood pressure by changing their lifestyle habits.

Objective

To the best of our knowledge, previous literature reviews have not incorporated a comprehensive list of studies on web-based dietary and physical activity–related intervention programs for patients with hypertension. Therefore, the objective of this scoping review was to provide a comprehensive view of the literature on the use of web-based dietary and physical activity–related intervention programs for patients with hypertension. This review sought to identify and summarize the methodologies, effectiveness, protocols, and theories, which enables us to have a better understanding of the web-based intervention programs that could influence patients' eventual health outcomes and improve existing clinical operations.

The researchers have carried out a scoping review of the studies published in the literature that were related to the digital and web-based dietary or physical activity intervention programs targeted for patients with hypertension. This review was based on the 5-point framework that was proposed by Arksey and O'Malley [13] and attempted to answer some questions as follows: (1) What were the focus areas or characteristics of the web-based learning programs? (2) What studies or interventions were based on the web-based intervention programs for managing hypertension? (3) What were the measurement outcomes for the web-based learning programs? (4) What was the distribution of studies that were focused on web-based intervention programs for supporting hypertensive disorders in the past 10 years?

Methods

Scoping Review Methodology

In this study, a scoping review methodology is chosen to acquire an overview of the extant literature on web-based dietary and physical activity–related intervention programs for patients with hypertension. A scoping review is a form of literature review that is useful for mapping applicable literature in a field of interest. The researchers carried out a scoping review based on the 5-step framework proposed by Arksey and O'Malley: (1) identifying the research objectives and search strategies; (2) identifying the relevant research studies; (3) study selection; (4) extracting and charting the data; and (5) collating, summarizing, and reporting all results [13].

A scoping review technique is very helpful for reviews that determine the range, extent, or nature of a specific topic. It is also used for classifying the main elements in a field or for identifying the gaps that exist in the related literature [14]. This review approach is also helpful if the study area is very broad or there is little or no information available about the topic. In such situations, a formal systematic review (which focuses on the study design and includes limiting selection criteria) limits the amount of information that can be acquired.

This review was conducted following the guidelines of the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for scoping review) [15]. PRISMA refers to the evidence-based minimal set of items to be reported for systematic reviews or meta-analyses. It includes a 27-item checklist along with a 4-phase flow diagram that helps all researchers report the retrieved data. The purpose of the PRISMA-ScR is to help readers develop a greater understanding of relevant terminology, core concepts, and key items to report for scoping review. Furthermore, it helps the researchers critically appraise all the previously published systematic reviews [16].

Search Strategy

The researchers carried out a search for the various web-based dietary and physical activity–related intervention programs for patients with hypertension in May 2020, with the help of a few comprehensive search databases. Table 1 presents the list of search terms that were used by the researchers. Since the objective of this study was to determine the web-based tools used in the health care sector, the researchers have focused their complete search on a few specific websites, such as PubMed, Web of Science, MEDLINE (Medical Literature Analysis and Retrieval System Online), ScienceDirect, Scopus, and Google Scholar. They used a combination of keywords, which included terms like web-based, diet, and physical activity. They set the search string with the help of an iterative process for including all the related keywords (Multimedia Appendix 1). They also studied the bibliography and the list of references included in the articles to determine the relevant studies. Furthermore, they manually searched for the keywords so that they could identify all the articles that they missed in the reference lists or databases.

Table 1. The search terms that were used by the researchers.

Keywords	Synonymous terms
Web-based	Mobile health OR internet OR mhealth OR digital OR eHealth OR web application OR internet-enabled interactive multimedia OR internet-supported online OR technology-assisted OR computer-assisted OR website-delivered OR computer-based OR computer-delivered OR internet-based OR interactive computer-based OR e-learning OR website OR digital medicines OR digital health technology
Nutrition intervention	Diet intervention
Physical activity	Exercises
Hypertension	Hypertensive OR high blood pressure

Selection Criteria

The researchers screened for the articles that were to be included in 2 stages: (1) title and abstract reviews and (2) complete article review. They used the scoping review technique for this purpose.

Eligibility and Exclusion Criteria

The aim was to include articles that describe web-based dietary and physical activity–related intervention programs for patients with hypertension. Studies were included if they (1) were an original paper published in a peer-reviewed journal (except review papers); (2) included website-based interventions for hypertension; (3) included diet, physical activity, or both as part of the intervention in the web-based intervention tools; (4) reported outcomes from an intervention using a website or digital health solution that involved use of the web for the same domain; (5) had adults diagnosed with hypertension as participants; (6) were published between January 2010 and March 2020; and (7) were written or translated in the English language.

The exclusion criteria were as follows: (1) studies targeting very specific groups of individuals (pregnant women, children, or athletes), (2) interventions only targeting health care providers, (3) studies written in other languages or not translated in English language, and (4) studies published in research avenues other than research journals. However, studies targeting people who are obese or diabetic were included as these diseases are directly linked to dietary habits and more prevalent in the general population.

Study Identification

The primary author (FNBA) carried out a literature search with the help of a search string (Multimedia Appendix 1). They screened the retrieved titles and the abstracts to eliminate the duplicate search results. They also eliminated all articles that were not related to the web-based dietary or physical activity–based intervention program.

Article Selection

Thereafter, 3 independent reviewers including the primary author (FNBA, MRAH, and SSB) reviewed the results and filtered the potentially related articles and the articles that were labelled as *maybe* by the primary author by using the above-mentioned inclusion and exclusion criteria. All discrepancies were resolved by consensus, and the researchers calculated the Cohen κ coefficient for measuring the interrater agreement (κ=0.76), which indicated a substantial agreement. Finally, the researchers acquired the full-text articles from their filtered list. Thereafter, 2 authors (FNBA and MRAH) reviewed all the acquired full-text articles and eliminated the articles that did not fulfil the inclusion and exclusion criteria.

Data Extraction

All data were extracted from the articles that were selected. These data have been described based on the (1) article characteristics (ie, which disease the study was based on and the study settings); (2) tool characteristics, that is, structure (medium of communication), function (additional features such as the viewable care plan), and communication paradigm (ie, one-to-many or one-one communication flow); (3) intended usage, context, and users included in the study; and (4)

evaluation of the study (ie, design, evaluation stages, and outcomes).

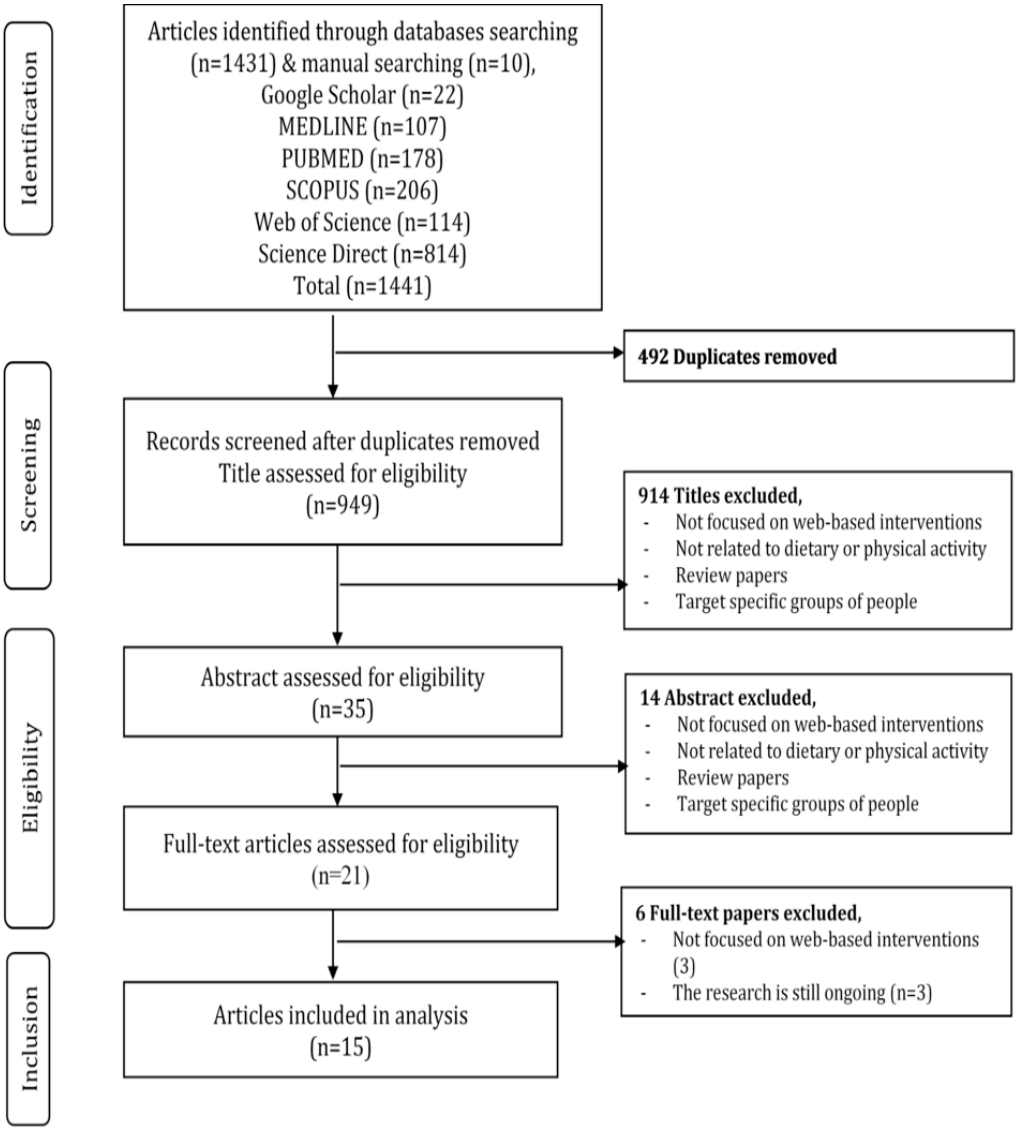
Results

Included Studies

After searching for relevant studies in the literature, the researchers retrieved 1441 studies from the 6 databases used: ScienceDirect (n=814), Scopus (n=206), PubMed (n=178), Web of Science (n=114), MEDLINE (n=107), and Google Scholar (n=22). After duplicates were eliminated, 949 results were obtained, out of which 914 studies were excluded based on their titles after the screening process.

Thereafter, the researchers screened the remaining 35 studies based on their titles and abstracts and downloaded 21 full-text articles. These articles were further reviewed based on the inclusion and exclusion criteria, which finally led to 15 articles (Figure 1). The 6 articles were excluded either because they did not include a web-based tool (n=3) or because research was still in progress (n=3). Lastly, the researchers extracted the data from the final 15 articles.

Figure 1. Flow diagram depicting the inclusion of studies that used web-based dietary and physical activity intervention programs for patients with hypertension.



Description of the Web-Based Tools Included in the Study

Article Characteristics

The earliest article that was published in the past decade was in 2011. Since then there was a significant increase in the number of studies that were published between 2011 and 2019, except in 2013 (Table 2). Of the 15 studies, most were conducted in the United States (5, 33%) and Canada (4, 27%) (Table 2). Many of these studies were carried out at primary care centers (11, 73%), which included the target adult population (15/15, 100%).

Characteristics of the Web-Based Tools

The web-based tools were characterized by their tool structures, functions, and communication paradigms. Table 3 presents these characteristics.

Table 2. Characteristics of the published articles (n=15).

Characteristics	n (%)
Publication year (N=15)	
2011	1 (7)
2012	2 (13)
2013	0 (0)
2014	1 (7)
2015	1 (7)
2016	2 (13)
2017	2 (13)
2018	3 (20)
2019	3 (20)
2020	0 (0)
Publication country of origin (N=15)	
England	1 (7)
United States	5 (33)
Canada	4 (27)
Spain	1 (7)
Korea	2 (13)
Italy	1 (7)
Taiwan	1 (7)
Unique study type (N=15)	
Original study	12 (80)
Protocol	1 (7)
Other: pilot study	2 (13)
Study context or setting of use (N=15)	
Primary care	11 (73)
Tertiary care outpatient clinics	1 (7)
Integrated health care organization	1 (7)
Workplace	1 (7)
Unknown	1 (7)
Population (N=15)	
Adult	15 (100)
Disease or clinical area of interest (N=18)[a]	
Cardiovascular disease	2 (11)
Hypertension	12 (67)
Metabolic syndrome	1 (6)
Diabetes	1 (6)
Hyperlipidemia	1 (6)
Obesity	1 (6)

[a]A few of the above studies assessed the tools in several contexts, for instance, to assess both mental health and diabetes in the patients.

Table 3. Tool characteristics, intended use, and users (n=15). The table classifies variables according to unique tools rather than individual studies as the unit of analysis.

Characteristics	n (%)
Structures	
Medium of communication or format[a]	
Website-based	15 (100)
Software-based connected by the internet	1 (7)
Telephone calls	1 (7)
Native application	2 (13)
Component of another platform (n=15)	
Patient portal	1 (7)
Electronic health record	2 (13)
Digital management program	1 (7)
Standalone	11 (73)
Functions	
Communication structure (n=15)	
One-way communication	5 (33)
Two-way communication	10 (67)
Type of communication (n=15)	
Unstructured communication (patient-provider free-form dialogue)	8 (53)
Structured communication (tailored feedback)	7 (47)
Tools with functions beyond patient-provider communication component[a]	
With 3 or more additional functions	4 (27)
Linked to a health record	3 (20)
Linked to appointment or scheduling	2 (13)
Linked to viewable care or treatment plan	2 (13)
Linked to prescription renewal	1 (7)
Linked to laboratory or test results	1 (7)
Linked to symptom diary or tracker	1 (7)
Linked to disease information or education	13 (87)
Linked to consultations	2 (13)
Communication paradigm	
Asynchronous	9 (60)
Of asynchronous tools, time-limited (response from a provider within a specified time window)	1 (7)
Synchronous	5 (33)
Both	1 (7)
Intended use and users	
Intended use of communication intervention[a]	
Lifestyle or behavior modification	13 (87)
Symptom reporting	1 (7)
Care planning	3 (20)
Medication adherence	6 (40)
Not specified	1 (7)

Characteristics	n (%)
Type of health care provider intended to use tool with patients or caregivers[a]	
Nurse	5 (33)
Physician	9 (60)
Pharmacist	2 (13)
Allied health professional	5 (33)
Health advisor	1 (7)
Not specified	1 (7)
Other	
Compensation to health care provider (n=15)	
Did not provide compensation	14 (93)
Did provide compensation	1 (7)
Tool access (n=15)	
Free through research participation	13 (87)
Prior registration required via website or service	2 (13)
URL available in article (n=15)	
Yes	4 (27)
No	11 (73)

[a]A few of the studies used multiple variables. For instance, with regard to the communication medium, one study used web-based tools and telephone calls.

Structures

A majority of the selected 15 web-based tools had a 2-way communication system between the patients and the providers (10, 67%). They used the website as the mode of communication. However, a few of these tools used multiple modes of communication, such as software-based systems that were connected by internet services (1, 7%), telephone calls (1, 7%), or native applications (2, 13%), which were developed as computer software or a smartphone app.

All the above-mentioned web-based tools were designed as educational or informational websites. However, only some of these web-based tools belonged to a different platform having multiple functions and features such as electronic health records (2, 13%), digital management programs (1, 7%), or a patient portal (1, 7%); 73% (11/15) of the tools were standalone communication tools.

Functions

The researchers identified 2 types of communication functions (ie, unstructured and structured communication). Many of these tools (8, 53%) included an unstructured communication system that allowed the providers and patients to engage in a free-form dialogue. Of the 15 web-based tools, 7 included a structured communication type, wherein the patients submitted their queries, which generated a response from their providers.

Furthermore, 87% of these tools (13/15) provided disease-related information and educated the people. In addition, a few of these tools offered different functions, that is, they were linked to the health records of the patients (3, 20%), appointments or schedules (2, 13%), treatment plans and a viewable care plan (2, 13%), consultation (2, 13%), prescription renewals (1, 7%), test or laboratory results (1, 7%), or a symptom tracker or diary (1, 7%).

Communication Paradigm

Most of these web-based tools (9, 60%) used an asynchronous communication system. Out of the different asynchronous tools, a single web-based tool showed a time limitation (7%). However, 5 of the 15 web-based tools used a synchronous communication system (33%), while 1 tool used both the communication types (7%).

Intended Use, Context, and Users

Based on the intended use of the various web-based tools, they were categorized into 4 different classes: behavior or lifestyle modifications (13, 87%), adherence to medicines (6, 40%), care planning (3, 20%), and reporting of symptoms (1, 7%). Out of the 15 web-based tools, 1 did not specify its intended use in the article, since the website only assessed the number of visitors who returned to the website.

A majority of the selected studies were conducted in a single critical field of study, that is, hypertension (12/15, 80%), while some papers used these web-based tools to evaluate multiple types of diseases such as cardiovascular diseases (2/18, 11%), diabetes (1/18, 6%), hyperlipidemia (1/18, 6%), metabolic syndromes (1/18, 6%), and obesity (1/18, 6%).

The health care workers who mainly used the web-based tools were physicians (9, 60%). In 5 of the selected studies, nurses and other allied health professionals like dietitians (33%) also used these tools. Of these studies, 13% included pharmacists (2/15), while 7% included health advisors. One study did not

specify the health care provider who used their tool. Out of the 15 selected studies, 1 study stated that the participants were offered monetary compensation for their participation.

Evaluation Characteristics

Study Design and Study Stage

Table 4 presents the evaluation of the characteristics displayed by the selected studies. 10 of the selected studies were randomized controlled trials (RCTs), while the remaining 5 were nonrandomized studies. Out of these nonrandomized studies, 2 included a quasi-experimental pretest versus posttest design, while 2 others were pilot studies and 1 included a website analysis. With regard to the stage of the study, based on the 2008 Medical Research Council Framework for Evaluation of Complex Interventions, 87% of the selected studies (13/15) were in their evaluation phase (10 RCTs and 3 nonrandomized studies). The remaining 2 nonrandomized studies were in their feasibility and piloting phases (13%).

The sample sizes of the RCT studies ranged between 106 and 9298 patients. These studies spanned for 4-12 months for follow-up. On the other hand, the sample sizes of the nonrandomized studies ranged between 56 and 803 patients for the quasi-experimental pretest versus posttest study design, while they ranged between 51 and 690 patients for the pilot studies.

Table 4. Assessment of the characteristics of unique studies (n=15). Unique studies were defined as the studies that led to several publications. These did not include editorials, protocols, or commentaries.

Study design and evaluation characteristics	Value
Randomized controlled trials (n=10)	
Communication component feature type, n	
Primary feature	5
Supplemental feature	5
Stage of study[a]	Evaluation
Types of results captured in each study[b], n	
Acceptability	2
Clinical	8
Psychological	1
Behavioral	5
Usage	1
Willingness	1
Cost-effectiveness	1
Sample size, median (IQR; range)	264 (165.75-369.25; 106-9298)
Length of follow-up (months), median (IQR; range)	12 (6.75-12; 4-12)
Nonrandomized studies (n=5)	N/A[a]
Quasi-experimental pretest vs posttest design (n=2)	
Communication component feature type, n	
Primary feature	1
Supplemental feature	1
Stage of study[a]	Evaluation
Types of results captured in each study[b], n	
Clinical	2
Sample size, median (IQR; range)	429.5 (242.75-616.25; 56-803)
Length of follow-up (months), median (IQR; range)	16 (12-20; 8-24)
Pilot study (n=2)	
Communication component feature type, n	
Primary feature	1
Supplemental feature	1
Stage of study[a]	Feasibility and piloting
Types of results captured in each study[b], n	
Usage	1
Usability	1
Clinical	1
Sample size, median (IQR; range)	370.5 (210.75-530.25; 51-690)
Length of follow-up (months), median (IQR; range)	6.25 (4.875-7.625; 3.5-9)
Website analysis (n=1)	
Communication component feature type, n	
Primary feature	0
Supplemental feature	0

Study design and evaluation characteristics	Value
Not specified	1
Stage of study[a]	Evaluation
Types of results captured in each study[b], n	
Usage	1
Sample size, n	38
Length of follow-up (months), n	6

[a]The definitions for the terms were derived from the 2008 Medical Research Council Framework for Evaluation of Complex Interventions.
[b]All results in the study have been included. The multiple outcomes are counted from the individual studies.

Study Outcomes

Table 4 presents the outcomes captured. A majority of the RCT studies (n=19 outcomes measured) focused on the clinical outcome, that is, a decrease in the blood pressure (8, 42%). Additionally, several studies investigated other outcomes such as behavioral changes (5, 26%), acceptability (2, 11%), cost-effectiveness (1, 5%), psychological effects (1, 5%), usage (1, 5%), and willingness (1, 5%). On the other hand, the nonrandomized studies (n=6 outcomes measured) investigated the clinical outcomes (3, 50%), usage (2, 33%), and usability (1, 17%) of the above web-based tools.

Discussion

Principal Findings

To our knowledge, this is the first study that has conducted a scoping review of published studies that used web-based tools for patients with hypertension. The researchers found 15 published studies that used web-based tools that could be helpful for patients with hypertension. These web-based tools were predominantly accessed from websites and mainly functioned as educational websites for diet, physical activity, or both. Few web-based tools provided multiple functions and features that enabled patient to access a patient portal, health records, or making appointments. Tools were used for behavior and lifestyle modifications, medication adherence, symptom reporting, and care planning.

Some of the tools enabled patients to communicate with health care providers such as physicians, nurses, and other allied health practitioners. Most tools were studied in the hypertension context, and some studies used them to evaluate multiple types of diseases including cardiovascular disease, diabetes, metabolic syndrome, hyperlipidemia, and obesity. More than half of the studies were RCTs that focused on clinical outcome evaluation, and several of them were observed behavioral changes. However, nonrandomized studies also examined impact on outcomes such as usage and usability.

Characteristics of the Web-Based Tools

The results indicated that the studies used the web-based tools on their websites as the mode of communication in order to improve the health of the patients, as they offered a lot of health-related information and attempted to educate the patients. A few web-based tools were included on a multifunctional platform for providing electronic health records (2, 13%), a patient portal (1, 7%), and a digital management program (1, 7%).

Additionally, a majority of these tools were used by health care providers such as physicians (9, 60%), allied health professionals (5, 33%), nurses (5, 33%), or pharmacists (2, 13%). The researchers noted that a multifunctional platform facilitates the patients and the health care practitioners as they are both able to simultaneously track the health records, objectives, and plans. Furthermore, this platform could help them derive all information and educate the patients.

Many studies stated that information technology (IT)–based products could be very helpful in the health care sector (health IT) as they allowed practitioners to easily communicate with their patients. They could also report some improvement in medication adherence or appointment compliance among the patients. These studies further showed that the use of health IT tools helped the patients connect with their health care providers. This improved the general communication between both the parties, encouraged the patients to become more autonomous, and empowered the patients to manage their personal health [17-19]. Furthermore, health IT offers benefits to the patients as these tools allowed the providers to offer a personalized health care plan to their patients.

Digital personalized health plans are a new, patient-targeted health care system that can improve the conventional health care system. It is seen to be a cost-effective technique that increases the coordination of health care and engagement rate of the patients. This, in turn, improves the outcome [20].

However, technological knowledge is usually accompanied by many challenges like a lack of digital literacy [21]. Based on the selected studies, the researchers noted that a few patients did not consent to participate in these studies as they were not well-educated or were elderly and lacked a digital education [22]. A different study determined the reasons why the patients refused to participate in such studies and noted that these patients lacked computer skills, had no internet facilities, did not understand that the digital, web-based tools could offer any benefits, or lacked time, but still participated in the conventional cardiac rehabilitation programs [23]. Harris et al stated that the people who showed a low health literacy rate and experienced health discrepancies were similar to the people who had a limited or no access to internet facilities, digital devices or showed a low digital literacy [24].

Hence, it is necessary to offer additional training to people before providing them with a health IT intervention tool to prevent any unintended risk to the users [25]. Stellefson et al stated that skill-building activities need to be integrated into the comprehensive patient education programs to allow the patients, particularly those with a low education level or limited knowledge, to identify the high-quality sources that provided web-based health information that was related to their diseases [26].

In this study, the researchers noted that many of the web-based tools (13, 87%) encouraged a behavioral or lifestyle modification, while some others included additional features like medication adherence (6, 40%), health care planning (3, 20%), or symptom reporting (1, 7%). Most of these lifestyle modifications focused on diet, physical activities, alcohol or smoking behavior, and physical activities.

Many theories were used in the selected studies like the health belief model, theory of planned behavior, social cognitive theory, transtheoretical model of behavior change, and cognitive-behavioral approach [27-31]. A few of the health care providers were also trained in motivational interviews [31,32]. The researchers noted that the implementation of these theories while designing informational or educational websites could help in improving the outcome.

Liu et al defined some behavioral goals for the participants depending on their readiness, which increased the efficacy of initiating any change and reviewing the cognitive and social behavioral skills for preventing relapse and maintaining adherence [27]. This was further supported by a different study wherein the participants were guided through the different readiness stages so that they adhered to the self-care behavior [28].

In their study, Sabooteh et al implemented the various stages of the behavior theory in the software-based and web-based educational intervention programs for improving the physical activity levels of the students [33]. According to the researchers, if the design of the intervention program was based on the preparation level of the patients to change and use novel educational tools, it showed a higher positive effect on their progress levels during different stages of their behavioral change levels.

Moeini et al assessed the efficiency of web-based depression improvement programs that were based on social cognitive therapy [34]. They observed that these intervention programs were able to improve the depression levels of the participants. This supported the fact that application of theories while designing the website intervention programs could improve the final result.

Here, the researchers have noted that different techniques were used by earlier studies for educating people or offering vital information. This information could be provided with the help of online information manuals, weblinks, or chat features (involving both video and text). This could be accessed through the patient portal or sending messages and emails which contain some weblinks to the patients.

The International Usability Professionals' Association stated that websites need to apply Section 508 to help every person, including those who face a difficulty hearing, seeing, or making accurate movements, so that they could use these links (US Department of Health and Human Services) [35]. Furthermore, the website content was an important aspect and must include very familiar words, while avoiding some terms that can confuse the users [35].

However, Lv et al held additional activities like cooking classes, healthy shopping tours, and educational webinars on the behavioral changes in order to improve patient engagement, rather than only distributing web-based educational handouts [29]. However, none of these engagement measures were related to the achievement of their blood pressure goals.

Similarly, Lussier et al compared the different approaches which included only the websites links and those that included web-based educational programs, followed by the workshop with an interactive nurse-facilitated small patient group [36]. They noted that the combined technique was not as effective in comparison to only the website-based approach, though they showed similar efficacy. This was attributed to the fact that the workshops were perceived as being outside of the routine care [37].

Evaluation Characteristics

Most of the studies included in this review focused on RCTs. The data from these studies indicated that many of the researchers compared website interventions with traditional interventions. Some of these studies assessed different aspects of the outcomes, including the clinical results. A few of them also measured several outcomes like usability, psychological results, usage, cost-effectiveness, and willingness.

However, there was not a lot of evidence regarding the usability of a website with the clinical outcomes, which highlighted the need to carry out website interventions. This step supported the use of a website-based educational program in the existing intervention techniques to improve existing clinical practices. A majority of these studies were conducted over a long period (ie, more than 12 months) with a follow-up and included 264 participants on average. However, positive results were noted within 3-4 months of assessment. For example, in their study, Kao et al observed that the patients who were offered a website-based intervention program showed a remarkable improvement in their clinical outcomes within 3 months of their assessment compared to the control group of patients [38].

In this study, the researchers noted that most of the outcomes showed a positive result after 12 months rather than the initial 3-4 months. This was supported by another study carried out by Cernvall et al wherein the researchers assessed the long-term efficacy of the internet-based guided self-help program (with a 12-month follow-up schedule) for the parents of children undergoing chemotherapy [39]. Their study indicated that the use of internet tools that offered psychological intervention was a very effective medium of delivery.

Murray et al investigated a web-based self-management program that was used by people suffering from type 2 diabetes for improving their glycemic control and decreasing diabetes-related

stress levels within 12 months [40]. A significant result was noted among the people using the program. Hence, the researchers stated that by offering web-based information to these patients, the health care providers could help them improve their overall health. This was regarded as an additional tool and could be used along with the existing self-management support techniques. This further showed that a website intervention program conducted for 12 months offered the best result.

Article Characteristics

Most of the above-mentioned studies were carried out in North America, especially in the United States (5, 33%) and Canada (4, 27%). A few studies were carried out in countries like the United Kingdom, Korea, Spain, Italy, and Taiwan. Though a website-based intervention technique was an effective tool for providing vital information to the hypertensive patients, the health care budgets were attributed as the primary reason why a majority of these studies were conducted in North America (n=9).

The results of this study also indicated that the highest number of papers on this topic were published between 2011 and 2012. Thereafter, very few reports were published between 2013 and 2015. This publication trend gradually increased between 2016 and 2019. One of the first reports that were published on this topic was carried out in the United States in 2011, followed by 2 papers that were published from Korea in 2012. Over the years, many studies that assessed the various website-based intervention tools that could be used by hypertensive patients were published from other countries.

Based on these observations, website-based education is an effective educational tool since it can be flexible and used by the patients anytime and anywhere. Laine et al noted that these website-based tools were very helpful as they offered the patients very vital information [41]. Furthermore, these tools also improved the mental state of the patients since they could control their physical health [42]. Thus, the patients could increase their knowledge and autonomously use and improve their health status.

Study Outcomes

The web-based tools were helpful for patients with hypertension. The selected studies showed positive results, which indicated that the patients with hypertension showed a significant improvement in their health when they incorporated the website-based tools into their regular schedule. Some of the studies measured the clinical outcomes of using these tools. One of the best clinical outcomes that were noted after the use of these tools was the marked decrease in the systolic and diastolic blood pressure of the patients. These were the main indicators that helped in assessing the “performance” of the website-based tools [43].

Here, the researchers have also noted that there was a significant decrease in the systolic blood pressure levels of the patients who used these tools for 12 months. Similar results were noted when traditional counselling was offered to the patients. Thus, it was concluded that the application of a web-based intervention tool in the clinical practices could improve the hypertension levels of the patients as these tools taught the patients to control their health status. Furthermore, the implementation of this technique encouraged them to execute long-term lifestyle changes and thus decrease their blood pressure levels [28].

Some of the selected studies also used usage and willingness, of the patients and the tools’ acceptability to the patients as important outcomes while measuring the effective nature of the website-based tools. They concluded that many people were willing to take part in a web-based educational program, based on the number of website visitors and returning visitors. By assessing these outcomes, the researchers determined the acceptability and the suitability of their website and found that it was important to effectively plan and then design a high-quality education website. It must include many multimedia elements, a higher accessibility level, and many interactive features [44].

Due to the advancement in the field of IT and the availability of online health information, website-based education could help in determining the health care behavior of patients and disease management [42]. Hence, the researchers in this study have proposed the need to determine the preferences of patients with regard to the features and design of educational websites for patients with hypertension. They aim to use patient-tailored information and design a website which offers interactivity, clear content presentation, content credibility, multimedia content and interpretability of the content, and other helpful features for educating the patients regarding their hypertension management.

However, this study displayed some limitations. The researchers used a scoping review technique, which identified the published articles at the expense of carrying out an in-depth analysis of the studies. The researchers aimed to carry out a comprehensive analysis using an extensive keyword search and search string for each database; however, they could have missed a few important studies, since there is a lack of standard terminologies that could be used in this area. The researchers also limited their review to include only studies published in English or translated in the English language. They could have missed some studies that were not translated to the English language or were not accessible from the databases used in the study, since many studies are published in countries that do not use English as their primary language.

Conclusions

Hypertension, or high blood pressure, is a trigger for many chronic diseases like cardiovascular diseases, diabetes, kidney failure, and strokes. Patients with hypertension can derive many benefits after using a web-based dietary and physical activity intervention program, which could help them make many lifestyle changes.

Here, the researchers selected 15 published studies that offered website-based intervention programs that included several interactive features for decreasing the blood pressure levels of the patients. These websites differed based on their content, usability, accessibility, and delivery method (eg, text, video, or audio). Before selecting a website program for their use, the patients considered many factors like ease of use, accessibility,

presence of additional features, and whether the offered information was supported by some evidence.

Hence, there is a lot of scope for improving some issues, such as the need for websites to be tested for their usability, acceptability, and resulting clinical outcomes. Based on these factors, it was determined whether the website could be implemented to decrease the blood pressure levels of the patients in a clinical setting. Use of these websites could decrease health care costs and the number of appointments since many people can access these websites at the same time, from different regions. Additionally, the efficiency and the quality of the care that is provided could be maintained or even improved.

Acknowledgments

This study was supported by a grant from the Ministry of Higher Education, Malaysia (reference number FRGS/1/2019/SS06/UiTM/03/6 and 600-IRMI/FRGS 5/3 [448/2019]). The authors would like to express their gratitude to the Ministry for this financial support and to the experts who gave a full commitment to this study. The data reported in this paper are those of the authors and are independent of the funding source.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search string.
[DOCX File , 18 KB - jmir_v23i3e22465_app1.docx]

References

1. Mills KT, Stefanescu A, He J. The global epidemiology of hypertension. Nat Rev Nephrol 2020 Apr;16(4):223-237. [doi: 10.1038/s41581-019-0244-2] [Medline: 32024986]
2. Mills KT, Bundy JD, Kelly TN, Reed JE, Kearney PM, Reynolds K, et al. Global Disparities of Hypertension Prevalence and Control: A Systematic Analysis of Population-Based Studies From 90 Countries. Circulation 2016 Aug 09;134(6):441-450 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.018912] [Medline: 27502908]
3. Singh S, Shankar R, Singh G. Prevalence and Associated Risk Factors of Hypertension: A Cross-Sectional Study in Urban Varanasi. Int J Hypertens 2017;2017:5491838 [FREE Full text] [doi: 10.1155/2017/5491838] [Medline: 29348933]
4. Pugh D, Gallacher PJ, Dhaun N. Management of Hypertension in Chronic Kidney Disease. Drugs 2019 Mar;79(4):365-379 [FREE Full text] [doi: 10.1007/s40265-019-1064-1] [Medline: 30758803]
5. Niiranen TJ, Kalesan B, Hamburg NM, Benjamin EJ, Mitchell GF, Vasan RS. Relative Contributions of Arterial Stiffness and Hypertension to Cardiovascular Disease: The Framingham Heart Study. J Am Heart Assoc 2016 Oct 26;5(11):e004271 [FREE Full text] [doi: 10.1161/JAHA.116.004271] [Medline: 27912210]
6. Petrie JR, Guzik TJ, Touyz RM. Diabetes, Hypertension, and Cardiovascular Disease: Clinical Insights and Vascular Mechanisms. Can J Cardiol 2018 May;34(5):575-584 [FREE Full text] [doi: 10.1016/j.cjca.2017.12.005] [Medline: 29459239]
7. Babaee Beigi MA, Zibaeenezhad MJ, Aghasadeghi K, Jokar A, Shekarforoush S, Khazraei H. The effect of educational programs on hypertension management. Int Cardiovasc Res J 2014 Sep;8(3):94-98 [FREE Full text] [Medline: 25177671]
8. Babaei-Sis M, Ranjbaran S, Mahmoodi H, Babazadeh T, Moradi F, Mirzaeian K. The Effect of Educational Intervention of Life Style Modification on Blood Pressure Control in Patients with Hypertension. J Edu Commun Health 2017 Apr 10;3(1):12-19. [doi: 10.21859/jech-03012]
9. Dzau V, Balatbat C. Future of Hypertension. Hypertension 2019 Sep;74(3):450-457. [doi: 10.1161/HYPERTENSIONAHA.119.13437] [Medline: 31352827]
10. Kitt J, Fox R, Tucker KL, McManus RJ. New Approaches in Hypertension Management: a Review of Current and Developing Technologies and Their Potential Impact on Hypertension Care. Curr Hypertens Rep 2019 Apr 25;21(6):44 [FREE Full text] [doi: 10.1007/s11906-019-0949-4] [Medline: 31025117]
11. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb;12(1):1-18 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
12. Fredericks S, Martorella G, Catallo C. A systematic review of web-based educational interventions. Clin Nurs Res 2015 Feb;24(1):91-113. [doi: 10.1177/1054773814522829] [Medline: 24571963]
13. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
14. Pham MT, Rajić A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res Synth Methods 2014 Dec;5(4):371-385 [FREE Full text] [doi: 10.1002/jrsm.1123] [Medline: 26052958]

15. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
16. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):1-8 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
17. Diamantidis CJ, Zuckerman M, Fink W, Hu P, Yang S, Fink JC. Usability of a CKD educational website targeted to patients and their family members. Clin J Am Soc Nephrol 2012 Oct;7(10):1553-1560 [FREE Full text] [doi: 10.2215/CJN.03690412] [Medline: 22798537]
18. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6] [Medline: 21071074]
19. Mbuagbaw L, Thabane L, Ongolo-Zogo P. Opening communication channels with people living with HIV using mobile phone text messaging: insights from the CAMPS trial. BMC Res Notes 2013 Apr 04;6:131 [FREE Full text] [doi: 10.1186/1756-0500-6-131] [Medline: 23557081]
20. Simmons LA, Drake CD, Gaudet TW, Snyderman R. Personalized Health Planning in Primary Care Settings. Fed Pract 2016 Jan;33(1):27-34 [FREE Full text] [Medline: 30766135]
21. Mohammadyari S, Singh H. Understanding the effect of e-learning on individual performance: The role of digital literacy. Computers & Education 2015 Mar;82:11-25. [doi: 10.1016/j.compedu.2014.10.025]
22. Green BB, Anderson ML, Ralston JD, Catz S, Fishman PA, Cook AJ. Patient ability and willingness to participate in a web-based intervention to improve hypertension control. J Med Internet Res 2011 Jan 20;13(1):e1 [FREE Full text] [doi: 10.2196/jmir.1625] [Medline: 21371993]
23. Clark RA, Tideman P, Tirimacco R, Wanguhu K, Poulsen V, Simpson P, et al. A pilot study of the feasibility of an Internet-based electronic Outpatient Cardiac Rehabilitation (eOCR) program in rural primary care. Heart Lung Circ 2013 May;22(5):352-359. [doi: 10.1016/j.hlc.2012.11.005] [Medline: 23294762]
24. Harris K, Jacobs G, Reeder J. Health Systems and Adult Basic Education: A Critical Partnership in Supporting Digital Health Literacy. Health Lit Res Pract 2019 Jul;3(3 Suppl):S33-S36 [FREE Full text] [doi: 10.3928/24748307-20190325-02] [Medline: 31773085]
25. Car J, Tan WS, Huang Z, Sloot P, Franklin BD. eHealth in the future of medications management: personalisation, monitoring and adherence. BMC Med 2017 Apr 05;15(1):73 [FREE Full text] [doi: 10.1186/s12916-017-0838-0] [Medline: 28376771]
26. Stellefson ML, Shuster JJ, Chaney BH, Paige SR, Alber JM, Chaney JD, et al. Web-based Health Information Seeking and eHealth Literacy among Patients Living with Chronic Obstructive Pulmonary Disease (COPD). Health Commun 2018 Dec;33(12):1410-1424 [FREE Full text] [doi: 10.1080/10410236.2017.1353868] [Medline: 28872905]
27. Liu S, Brooks D, Thomas SG, Eysenbach G, Nolan RP. Effectiveness of User- and Expert-Driven Web-based Hypertension Programs: an RCT. Am J Prev Med 2018 Dec;54(4):576-583. [doi: 10.1016/j.amepre.2018.01.009] [Medline: 29456025]
28. Liu S, Tanaka R, Barr S, Nolan RP. Effects of self-guided e-counseling on health behaviors and blood pressure: Results of a randomized trial. Patient Educ Couns 2020 Mar;103(3):635-641. [doi: 10.1016/j.pec.2019.10.007] [Medline: 31669047]
29. Lv N, Xiao L, Simmons ML, Rosas LG, Chan A, Entwistle M. Personalized Hypertension Management Using Patient-Generated Health Data Integrated With Electronic Health Records (EMPOWER-H): Six-Month Pre-Post Study. J Med Internet Res 2017 Sep 19;19(9):1-16 [FREE Full text] [doi: 10.2196/jmir.7831] [Medline: 28928111]
30. Mensorio MS, Cebolla-Martí A, Rodilla E, Palomar G, Lisón JF, Botella C, et al. Analysis of the efficacy of an internet-based self-administered intervention ("Living Better") to promote healthy habits in a population with obesity and hypertension: An exploratory randomized controlled trial. Int J Med Inform 2019 Apr;124:13-23. [doi: 10.1016/j.ijmedinf.2018.12.007] [Medline: 30784422]
31. Nolan RP, Feldman R, Dawes M, Kaczorowski J, Lynn H, Barr SI, et al. Randomized Controlled Trial of E-Counseling for Hypertension: REACH. Circ Cardiovasc Qual Outcomes 2018 Jul;11(7):1-10. [doi: 10.1161/CIRCOUTCOMES.117.004420] [Medline: 30006474]
32. Salisbury C, O'Cathain A, Thomas C, Edwards L, Gaunt D, Dixon P, et al. Telehealth for patients at high risk of cardiovascular disease: pragmatic randomised controlled trial. BMJ 2016 Jun 01;353:i2647 [FREE Full text] [doi: 10.1136/bmj.i2647] [Medline: 27252245]
33. Sabooteh S, Feizi A, Shekarchizadeh P, Shahnazi H, Mostafavi F. Effect of web-based and software-based educational intervention on stages of behavior change of students' physical activity. J Educ Health Promot 2020;9:123 [FREE Full text] [doi: 10.4103/jehp.jehp_645_19] [Medline: 32642479]
34. Moeini B, Bashirian S, Soltanian A, Ghaleiha A, Taheri M. Examining the Effectiveness of a Web-Based Intervention for Depressive Symptoms in Female Adolescents: Applying Social Cognitive Theory. J Res Health Sci 2019 Aug 19;19(3):1-8 [FREE Full text] [Medline: 31586376]
35. Yadrich DM, Fitzgerald SA, Werkowitch M, Smith CE. Creating patient and family education web sites: assuring accessibility and usability standards. Comput Inform Nurs 2012 Jan;30(1):46-54 [FREE Full text] [doi: 10.1097/NCN.0b013e3182343eac] [Medline: 22024970]

36. Lussier M, Richard C, Glaser E, Roberge D. The impact of a primary care e-communication intervention on the participation of chronic disease patients who had not reached guideline suggested treatment goals. Patient Educ Couns 2016 Apr;99(4):530-541. [doi: 10.1016/j.pec.2015.11.007] [Medline: 26657041]
37. Krist AH, Woolf SH, Bello GA, Sabo RT, Longo DR, Kashiri P, et al. Engaging primary care patients to use a patient-centered personal health record. Ann Fam Med 2014;12(5):418-426 [FREE Full text] [doi: 10.1370/afm.1691] [Medline: 25354405]
38. Kao C, Chen T, Cheng S, Lin W, Chang Y. A Web-Based Self-Titration Program to Control Blood Pressure in Patients With Primary Hypertension: Randomized Controlled Trial. J Med Internet Res 2019 Dec 05;21(12):e15836 [FREE Full text] [doi: 10.2196/15836] [Medline: 31804186]
39. Cernvall M, Carlbring P, Wikman A, Ljungman L, Ljungman G, von Essen L. Twelve-Month Follow-Up of a Randomized Controlled Trial of Internet-Based Guided Self-Help for Parents of Children on Cancer Treatment. J Med Internet Res 2017 Jul 27;19(7):e273 [FREE Full text] [doi: 10.2196/jmir.6852] [Medline: 28751300]
40. Murray E, Sweeting M, Dack C, Pal K, Modrow K, Hudda M, et al. Web-based self-management support for people with type 2 diabetes (HeLP-Diabetes): randomised controlled trial in English primary care. BMJ Open 2017 Sep 27;7(9):e016009 [FREE Full text] [doi: 10.1136/bmjopen-2017-016009] [Medline: 28954789]
41. Laine A, Välimäki M, Löyttyniemi E, Pekurinen V, Marttunen M, Anttila M. The Impact of a Web-Based Course Concerning Patient Education for Mental Health Care Professionals: Quasi-Experimental Study. J Med Internet Res 2019 Mar 01;21(3):e11198 [FREE Full text] [doi: 10.2196/11198] [Medline: 30821697]
42. Win KT, Hassan NM, Bonney A, Iverson D. Benefits of online health education: perception from consumers and health professionals. J Med Syst 2015 Mar 11;39(3):27-1-27-8. [doi: 10.1007/s10916-015-0224-4] [Medline: 25666928]
43. Lennon SL, DellaValle DM, Rodder SG, Prest M, Sinley RC, Hoy MK, et al. 2015 Evidence Analysis Library Evidence-Based Nutrition Practice Guideline for the Management of Hypertension in Adults. J Acad Nutr Diet 2017 Sep;117(9):1445-1458.e17. [doi: 10.1016/j.jand.2017.04.008] [Medline: 28578899]
44. Emre Dinc EB. Web-based education and accessibility. International Journal of Technology in Education and Science (IJTES) 2017;1(1):29-35.

Abbreviations

IT: information technology
LMICs: low- and middle-income countries
PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 13.07.20; peer-reviewed by K Dasgupta, P Banik; comments to author 30.10.20; revised version received 13.11.20; accepted 16.01.21; published 15.03.21.

Please cite as:
Baderol Allam FN, Ab Hamid MR, Buhari SS, Md Noor H
Web-Based Dietary and Physical Activity Intervention Programs for Patients With Hypertension: Scoping Review
J Med Internet Res 2021;23(3):e22465
URL: https://www.jmir.org/2021/3/e22465
doi:10.2196/22465
PMID:33720036

Original Paper

Developing a Web-Based Shared Decision-Making Tool for Fertility Preservation Among Reproductive-Age Women With Breast Cancer: An Action Research Approach

Ling-Ming Tseng[1,2], MD; Pei-Ju Lien[1], BA; Chen-Yu Huang[3], MD, PhD; Yi-Fang Tsai[1,2], MD, PhD; Ta-Chung Chao[1,2,4], MD, PhD; Sheng-Miauh Huang[5], PhD

[1]Comprehensive Breast Health Center, Department of Surgery, Taipei Veterans General Hospital, Taipei, Taiwan

[2]School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

[3]Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynecology, Taipei Veterans General Hospital, Taipei, Taiwan

[4]Division of Medical Oncology, Department of Oncology, Taipei Veterans General Hospital, Taipei, Taiwan

[5]MacKay Medical College, Department of Nursing, New Taipei City, Taiwan

Corresponding Author:
Sheng-Miauh Huang, PhD
MacKay Medical College
Department of Nursing
No. 46, Section 3, Zhongzheng Road
Sanzhi District
New Taipei City, 252
Taiwan
Phone: 886 2 26360303 ext 1316
Email: r910862@yahoo.com.tw

Abstract

Background: The pregnancy rate after cancer treatment for female survivors is lower than that of the general population. Future infertility is a significant concern for patients with breast cancer and is associated with a poor quality of life. Reproductive-age patients with breast cancer have safe options when choosing a type of fertility preservation method to be applied. Better information and support resources aimed at women to support their decision making are needed.

Objective: The objective of this study was to develop a web-based shared decision-making tool for helping patients with breast cancer make decisions on fertility preservation.

Methods: We used the action research cycle of observing, reflecting, planning, and acting to develop a web-based shared decision-making tool. The following four phrases were applied: (1) observe and reflect—collect and analyze the decision-making experiences of patients and health care providers; (2) reflect and plan—apply the initial results to create a paper design and modify the content; (3) plan and act—brainstorm about the web pages and modify the content; (4) act and observe—evaluate the effectiveness and refine the website's shared decision-making tool. Interviews, group meetings, and constant dialogue were conducted between the various participants at each step. Effectiveness was evaluated using the Preparation for Decision-Making scale.

Results: Five major parts were developed with the use of the action research approach. The Introduction (part 1) describes the severity of cancer treatment and infertility. Options (part 2) provides the knowledge of fertility preservation. The shared decision-making tool was designed as a step-by-step process (part 3) that involves the comparison of options, patient values, and preferences; their knowledge regarding infertility and options; and reaching a collective decision. Resources (part 4) provides information on the hospitals that provide such services, and References (part 5) lists all the literature cited in the website. The results show the web-based shared decision-making meets both patients' and health providers' needs and helps reproductive-age patients with breast cancer make decisions about fertility preservation.

Conclusions: We have created the first web-based shared decision-making tool for making fertility preservation decisions in Taiwan. We believe female patients of reproductive age will find the tool useful and its use will become widespread, which should increase patient autonomy and improve communication about fertility preservation with clinicians.

Trial Registration: Clinicaltrials.gov NCT04602910; https://clinicaltrials.gov/ct2/show/NCT04602910

(J Med Internet Res 2021;23(3):e24926) doi:10.2196/24926

KEYWORDS

breast cancer; shared decision making; website; action research; fertility preservation

Introduction

Breast cancer is a common oncologic disease worldwide. In Taiwan, breast cancer has the highest incidence of all cancers in the female population, and the incidence of breast cancer in women of childbearing age is increasing [1]. Approximately 12% to 19% of patients are affected when they are of reproductive age [1-3]. Fertility and infertility are important concerns among reproductive-age women with breast cancer [4,5]. Infertility following cancer treatment has a recognized negative impact on the quality of life [6]. The National Cancer Institute (Taiwan) defined fertility preservation as a type of procedure used to help a person' retain the ability to have children [7]. The key point is that reproductive-age patients have more than one fertility preservation choice. Unfortunately, most patients do not have enough information to make an informed decision prior to cancer treatment. For instance, ovarian stimulation in patients with early-stage breast cancer is safe in the long term [3]. Women diagnosed with cancer who have eggs or embryos cryopreserved before anticancer treatment have good chances for successful assisted reproductive technology performance and good perinatal outcomes [8-10]. In addition, the efficacy and safety of temporary ovarian suppression with gonadotropin-releasing hormone agonist during chemotherapy might be an available option to reduce the likelihood of chemotherapy-induced primary ovarian insufficiency and to improve future fertility in premenopausal patients with early breast cancer [11]. It is worth noting that there is consensus in medical society that the long-term survival rate of the disease is not lower for any of the three types of fertility preservation [8-11]. Thus, if age, reproductive function, and the cancer stage allow, patients can choose their preferred fertility preservation method.

A previous study [12] indicated patients' decisions about fertility are multidimensional. The risk perception of pregnancy among patients with breast cancer after treatment focused on "reaching the balance of life [12]." Women treated for breast cancer applied risk-benefit perceptions to decide whether to become pregnant [12]. There are several factors related to the ability to make well-informed decisions regarding fertility preservation. First, patients with a new diagnosis of breast cancer often experience negative emotions, such as anxiety, depression, and uncertainty, due to their perceptions of life-threatening cancer [13-15]. Second, women are expected to have biological children because of the Chinese cultural belief in lineal descent from one's ancestors, which is deeply rooted among the Taiwanese people. The assignment of great importance to parenthood was directly associated with higher depression symptoms in reproductive-age women with breast cancer [16].

Additionally, some health care providers might keep silent about discussing fertility because their primary concern is getting their patients to be cancer free [17]. Patients with breast cancer lacked easily available knowledge about infertility and underestimated the possibility of infertility [18]. This requirement for information is mostly unmet. Patients do not fully understand the impact of cancer treatment and the possibility of infertility because of treatment. Finally, the space and environment of the outpatient clinic are not conducive to the discussion of private issues. Fertility concerns are not encouraged to be brought to the physicians' or nurses' attention before cancer treatment. Although women want to request further information regarding future fertility, they may hesitate even when talking to their physicians, disclosing their opinion, or showing their personal feelings.

The first Patient Autonomy Act was passed through the Legislative Yuan in Taiwan in 2015. Now, it is vital to ensure that patients have the right to know, choose, or refuse medical care, and in addition to being informed, patients can choose their medical options. All of the above can be achieved via shared decision making (SDM) [19-22]. The Joint Commission of Taiwan, which is an organization established in 1999 with funding from the Ministry of Health and Welfare, Taiwan Hospital Association, promotes the development of SDM [23]. Difficulty accessing decision aids or limited patient comprehension were common barriers to realizing shared decision making [24]. Decision aids that improve decision-related outcomes for many breast cancer treatment decisions, including surgery, radiotherapy, and endocrine and chemotherapy, are available [25]. Fertility preservation with decision aids is well-developed in some economically high-income countries, such as Australia [26], Canada [27], Netherlands [28], and the United States [29]. Therefore, we aimed to develop a web-based SDM tool for fertility preservation for patients with breast cancer. The web-based decision aid was expected to provide medical information, help the patient explore and compare treatment options, assess the patient's values and preferences, and reach a collective decision on fertility preservation.

Methods

Research Design

We applied an action research design, which is defined as an approach that involves collaboration to develop a process through knowledge-building and social change [30]. Learning by doing with the participants, including the patients, is the heart of health care action research. The action research approach is particularly relevant when treating patients with chronic diseases and complex care needs [31,32]. Previous investigators have applied action research to build web-based comics [33] or decision-aid websites [34] about breast cancer surgery in Taiwan. However, few study the decision-making and research processes for oncofertility in Asia. Therefore, the ultimate goal of this study was to find supporting evidence of the benefits to patients and families of a web-based platform that assists patients with breast cancer in the SDM for fertility preservation.

Project Team

The SDM team was multidisciplinary, comprising a senior researcher, two surgeons, an oncologist, a gynecologist, an advanced nurse specialist, and an information technology (IT) engineer. The surgeon, oncologist, and gynecologist were responsible for gathering and reviewing existing fertility preservation literature to create easy-to-understand formats and simple graphic renderings of patients with breast cancer. The researcher and nurse specialist interviewed women and health providers to explore their experiences regarding the decision-making process. The IT engineer was responsible for informatics work related to website page design, program coding, and hosting of the website. The members involved in breast care had at least 10 years of experience. The IT engineer was a senior website designer with 5 years of work experience. The team had regular face-to-face and line meetings to prepare for and reflect on the development of the oncofertility SDM program and the community of fertility preservation practice. The SDM team was the partnership in the sense of doing together and deciding together during the research phase. The roles of the members were alternately those of initiator, educator, facilitator, coach, and finally, coauthors.

Action Research Process

A participatory action research methodology was used to facilitate the development of the SDM program for the multidisciplinary care team at the same time that a web-based SDM of fertility preservation practice was being established. The action research cycle of observing, reflecting, planning, and acting was applied in this study (Figure 1).

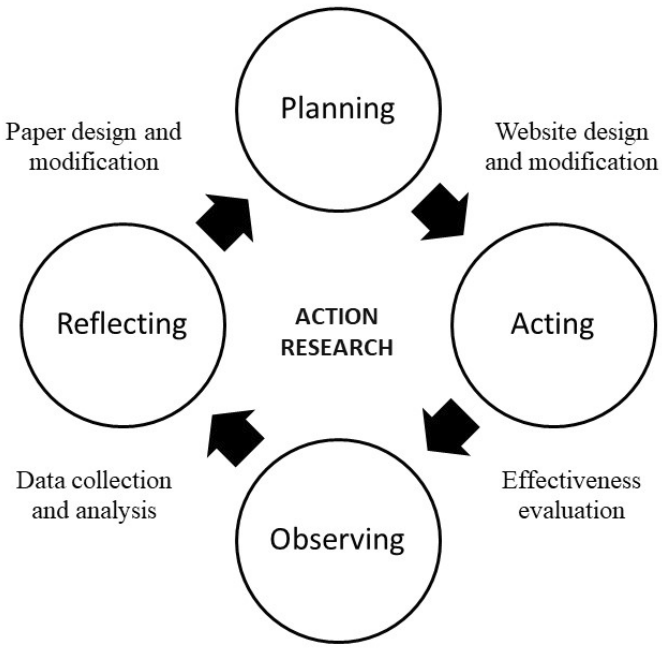
Because the SDM involved both health care providers and patients, we collected data from both views at the phases of observing and reflecting. Health care providers were recruited to interview and ask patients about their experiences regarding fertility. We also recruited breast cancer survivors to explore the decision-making experience of pregnancy. The reflections were viewed as a continuous dialogue among team members during the SDM program. The results of data collection and team self-evaluation of our positionality were analyzed, with the explicit recognition that this position may affect the SDM process and outcome. The evaluations were interdependent and partly based on the project team members' interpersonal, social, and institutional contexts.

During the phases of reflecting and planning, all major team members not only reviewed the health issues and evidence-based literature related to fertility preservation in patients with cancer but also discussed the SDM content based on the findings from the observe and reflect phases. Three patients were invited to participate and provide their suggestions. The initial paper-designed content of SDM for fertility preservation was formed by consensus after 6 meetings.

During the phases of planning and acting, the initial website structure was developed according to the 5 steps of SDM. Each part of the SDM was assessed for comprehensibility and usability by patients. Health care providers were asked to assess the acceptability of the SDM content. A 5-point Likert scale was used for patient and health care provider ratings for each item. If the score was less than 3, then we modified the website content based on the user feedback from patients and health care providers.

During the phase of acting, we evaluated the effectiveness of the Preparation for Decision-Making (prepDM) scale developed by Bennett et al [35]. The Joint Commission of Taiwan also suggested the effects of SDM be evaluated in patients and health care providers. Both patients and health care providers were invited to answer 10 questions by responding with a 5-point Likert scale rating (from 1 to 5). A higher score indicated greater agree with the effect.

Figure 1. The process of action research in developing web-based shared decision-making regarding fertility preservation.



Setting

The setting was the comprehensive breast health center at Taipei Veterans General Hospital in Taipei, Taiwan. The center integrates multidisciplinary care to provide comprehensive services, including diagnosis and treatment, to achieve the purpose of patient-centric care. The health care providers included surgeons, gynecologists, Chinese medicine doctors, nurses, psychological consultants, social workers, and nutritionists. The web-based SDM for fertility preservation was linked to the hospital website.

Analysis

In-depth interviews were done to collect and gain deep insights into the views and needs of the participants. Data collected through the face-to-face interviews were transcribed from audiorecordings to written transcripts. The content analysis method was used to analyze the data. We checked whether the categories were stable and provided sufficient depth, then multiple strategies were applied to ensure trustworthiness. The discussion and decision at the group meetings were also recorded. The effectiveness was evaluated using the prepDM scale developed by Bennett et al [35]. Statistical analyses were performed using the Predictive Analytics Suite workstation (version 18.0; IBM Corp). Individual variables were examined by percentages, means, and standard deviations.

Results

Results of the Observe and Reflect Phase

Cancer survivors (n=16) were recruited to explore the decision-making experience of determining whether to have a pregnancy (from 2015 through 2016). The result showed decision making regarding fertility among women with cancer was affected by preexisting needs for children before treatment and their experience during treatment. Underestimating the possibility of infertility and a lack of knowledge of fertility preservation might cause the woman to regret her decision. (Participant names are pseudonyms.)

> *If I had known in advance, I could have frozen my eggs. I mean, normal eggs are taken before chemotherapy. Because it [chemotherapy] has damaged my eggs now, I can't afford another invasive examination or treatment.* [Bella]

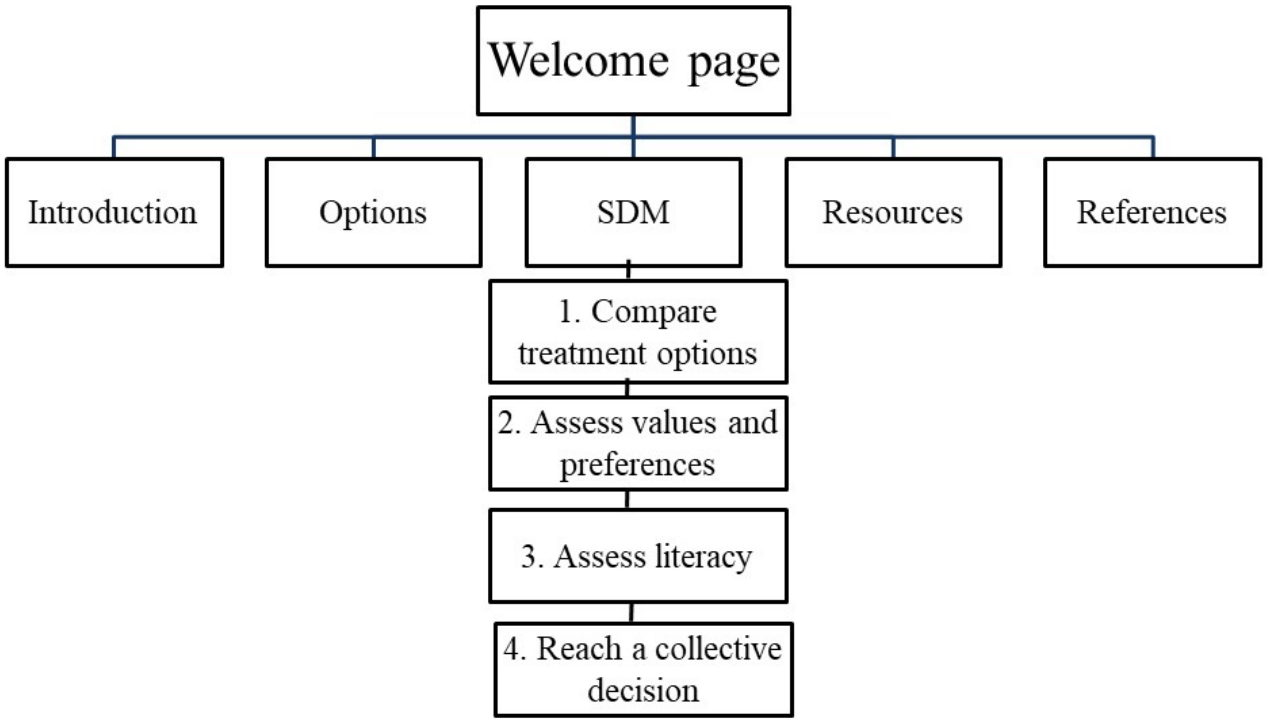
Health care providers (n=16), including nurses (n=8), surgeons (n=3), gynecologists (n=2), Chinese medicine doctor (n=1), psychology consultant (n=1), and social worker (n=1), were also recruited to interview and ask the women about their experiences regarding fertility (from 2017 through 2018). Although most of the health care providers were concerned about the severity of the cancer and the urgency of treatment, they agreed to satisfy the women's needs and respect their choice. However, there are some barriers to be solved, such as communication blocks among the multidisciplinary team, lack of initial screening, and lack of evidence-based information regarding oncofertility.

> *I think the empirical literature may exist, but nobody has assembled it into a simple language that the patient can understand. All we know is that it requires more manpower and material resources...* [Jeff]

Results of the Reflect and Plan Phase

After collecting the data and performing a content a qualitative content analysis, we discussed strategies to overcome the barriers. The SDM process was expected to progress step by step, according to the five essential steps of SDM developed by the Agency for Healthcare Research and Quality [36]. Reproductive-age women had insufficient knowledge regarding infertility and the types of fertility preservation. More than 96% of Taiwanese ever used the internet to search for information [37]. Hence, all project members focused the information they provided on the website on cancer treatment, infertility, and the types of fertility preservation necessary before embarking on the SDM steps. Therefore, the introduction involved the description of the severity of the cancer and its treatment and the possibility of infertility. Then, the options regarding fertility preservation were designed. The SDM part included providing comparisons of treatment and fertility preservation options, assessing each patient's values and preferences (such as cost or safety), assessing their degree of knowledge regarding infertility and fertility preservation options, and reaching a collective decision on the best option. We also linked the organizations that provide fertility preservation services in a resource section [38]. Appropriate literature was also provided in the reference section. The website structure was designed, as shown in Figure 2. All team members were assigned to fulfill this information need for the patients. The content provided in each section was reached by consensus in group meetings.

Figure 2. Website structure. SDM: shared decision making.



Results of the Plan and Act Phase

Upon entering the welcome page, patients are shown a clear title; for example, "I am a patient with breast cancer. How will I fulfill my dream of having a child? (Chinese: 我是乳癌病人，怎麼圓生子夢?)." Our webpage follows a structure where the five main parts with their corresponding submenus are displayed as the user clicks on them (Multimedia Appendix 1). The five main webpage subtitles are Introduction, Options, SDM, Resources, and Reference Links. A 5-point Likert scale was used for patient and care provider ratings. Patients (n=13) rated the web-based SDM for comprehensibility (mean 3.9, SD 0.9) and usability (mean 3.9, SD 1.1). Health care providers (n=9) were also asked to assess the acceptability (mean 4.3, SD 0.8). If the score was less than 3, we modified the website content based on the user feedback from patients and health care providers. For example, patients mentioned that pharmaceutical and scientific names or brand names (such as gonadotropin-releasing hormone agonist or Zoladex) are too hard to understand. Patients said it is easier to understand simpler terms like "medicine postponing menstruation" (Chinese: 注射停經針). When patients were tested on their knowledge regarding fertility preservation, some still misunderstood and felt frustrated. Hence, we provide links to help answer questions and clarify their knowledge. When all questions are answered correctly, the user can enter the next stages (Multimedia Appendix 2; Multimedia Appendix 3; Multimedia Appendix 4).

The Action Script technology for the front-end framework was Vue.js, and the back-end framework was Laravel 5 (requiring PHP 7.2). The operating system was Amazon Linux 2, while the network server was Apache 2.4. AWS EC2 hosted the web-based machine. The database was MySQL 5.7 (Amazon RDS).

Results of the Act and Observe Phase

We used a multifaceted approach to evaluate the website after the completion of the preliminary website. The SDM effectiveness of the website was evaluated using the SDM questionnaire developed by The Joint Commission of Taiwan. The participants in the evaluation consisted of 14 reproductive-age patients with breast cancer (n=14, prepDM: mean 4.1, SD 0.8) and 11 hospital staff (n=11, prepDM: mean 4.2, SD 0.7) (Table 1). Both patients and staff had the perception that the web-based SDM was effective for helping reproductive-age patients with breast cancer make fertility preservation decisions. Based on feedback from the patients and hospital care providers, we refined the options because some patients had hesitation about the website. The option "can't decide right now" was added, and qualitative and quantitative items regarding further needs were added to the survey (Multimedia Appendix 5).

Table 1. Results of the Preparation for Decision-Making Survey scale among health care providers and patients with breast cancer.

Item	Health providers (n=11), mean (SD)	Patients (n=14), mean (SD)
1. Recognize decision needs to be made	4.1 (0.8)	4.1 (0.9)
2. Prepare my patient/you to make a better decision	4.2 (0.9)	4.2 (1.0)
3. Think about pros and cons of each option	4.2 (0.8)	4.2 (0.8)
4. Help my patient/you think about which pros and cons are most important	4.3 (0.6)	4.3 (0.8)
5. Know that decision depends on what matters most to my patient/you	4.1 (0.8)	4.2 (0.8)
6. Organize my patient's/your thoughts about decision	4.2 (0.8)	3.9 (1.0)
7. Think about how involved my patient/you want to be in the decision	4.4 (0.8)	3.9 (0.9)
8. Help my patient/you identify questions you want to ask	4.4 (0.7)	4.1 (0.9)
9. Talk to patient's/your doctor about what matters most	4.3 (1.0)	4.1 (1.0)
10. Promote health literacy/prepare you for a follow-up visit with your doctor	4.3 (0.8)	3.9 (1.2)
Average	4.2 (0.7)	4.1 (0.8)

Discussion

Principal Findings

We used action research to create a website for the SDM model for fertility preservation among young women with breast cancer [39]. The approach consisted of four phases based on the principles of SDM. Previous SDM-related literature has highlighted the importance of the principles of informed consent and patient autonomy [19-22]. When women make a meaningful decision regarding their choice of fertility preservation, their degree of medical literacy, preferences, and a close relationship (mutual communication and understanding) between themselves and health care providers are often key challenges. Sometimes, a conflict happens between what the patients want and what the clinicians deem necessary in terms of the website during the four phases. We not only collected qualitative and quantitative data but brainstormed acceptable content for both health care providers and patients. This work implied that the SDM for fertility preservation, as guided by action research, would bridge the gap between the patients and clinicians and caters more toward the feelings and needs of the patients' minds and bodies than previous approaches. The findings were consistent with those of previous studies [33,34] for which decision-aid websites for breast cancer surgery or comics for newly diagnosed breast cancer in women were developed. Compared with the development of medical information websites, the development of web-based decision aids using action research may be more systematic and humanistic because they account for both the patients' and health care providers' perspectives and needs. When considering the participants' corresponding culture and the region in which they live, our study serves as a reference for describing how to develop decision support tools for women with breast cancer and fertility needs.

Limited health literacy is associated with poorer health outcomes and may make it difficult for cancer patients to participate in the SDM process [23]. Increasing health literacy has the potential to increase health care engagement and, subsequently, to increase the use of SDM [40]. Improving women's health literacy is a necessary strategy to promote informed consent and SDM. In Taiwan, explaining the disease condition and its related treatment is the responsibility of the attending physician. Physicians explain the disease and related treatments in colloquial language, which is a key factor in promoting patients' perceived involvement in SDM [41]. Before the SDM process, our Introduction and Options website pages provided familiar text and pictures for physicians to explain the consequences of infertility and the methods of fertility preservation. It was also easy for patients to go back and reread the pages for better understanding. Previous scholars mentioned three major SDM steps, which included exploring care or treatment options and their risks and benefits, discussing choices available, and reaching a decision about care or treatment, together with their health and social care professionals [21,42]. In our study, we added the literacy assessment before reaching a collective decision. The assessment enabled us to make sure our participants had enough knowledge regarding infertility and fertility preservation before making a decision. The impact of the health literacy intervention on decision regret and psychological changes merits future research.

Limitations

There were several challenges during the development of the website. The patient's psychological burden and fear increased with the severity of the cancer stage. Despite this situation, we did not provide different options for fertility preservation for patients with severer cancer stages because of insufficient evidenced-based literature. Evidence-related survival rates and recurrence rates among women with fertility preservation require more research.

With advancements in reproductive technologies, the evidence will be updated, and new options may appear periodically when the website is updated. The action research cycle provides an opportunity to modify and confirm this repeatedly, as it is quite a repetitive and lengthy process for the professionals involved.

The SDM webpages were limited to the Chinese version. Hence, we could not explore the advantages and disadvantages related to non-Chinese and non-English literature and websites. The web-based SDM regarding fertility preservation was only

designed for women with breast cancer. Nonetheless, this study can serve as a foundation and reference for health care providers who are interested in future planning, promotion, operation, and long-term management of websites for patients with cancers other than breast cancer. The SDM website was hosted by the hospital and depended on internal technical support. All researchers in our study are responsible for ensuring a responsive-interactive website. Considering information security, patients need to provide some personal information to enter the welcome page. If the patient uses the website alone and replies to the questions on the webpage, a time delay in getting a doctor's response may increase the patient's anxiety because patients and clinicians are encouraged to visit the website together so that the doctor can answer any questions. We only evaluated the effectiveness of web-based SDM using the prepDM scale. Using the full-fledged website evaluation heuristics and exploring the long-term consequences, such as decision regret and quality of life among reproductive-age women with breast cancer, merit more study in the future.

Conclusions

With the help of the study results, we built a web-based decision-aid tool for helping reproductive-age women with breast cancer make decisions on fertility preservation. The action research provided a good structure to guide cooperation involving patients, clinicians, nurses, academic researchers, and IT engineers. The research results helped our team manage effectively and shortened the distance between theory and practice. It also helped participants who are facing the dilemma of choosing a method of fertility preservation. Based on the importance of informed consent, our website not only provided knowledge of infertility and the methods of fertility preservation but allowed us to design and test this information. It helps ensure that our participants make fertility preservation decisions with knowledge and understanding. Based on our findings, we conclude that this SDM website can indeed help patients with early-stage breast cancer make more informed decisions regarding the type of fertility preservation they would prefer to undergo. Longitudinal studies to follow up on the changes in psychological condition and subsequent pregnancy rates are needed in the future.

Acknowledgments

The *observe and reflect* phase was supported by the Ministry of Science and Technology, Taiwan (MOST 103-2314-B-715-005-MY3; MOST 106-2314-B-715-004 -MY2). The *plan and act* phase was supported by Taipei Veterans General Hospital, Taiwan (V108E-005-1) and the Ministry of Science and Technology, Taiwan (MOST 109-2314-B-715-004). Melissa Lee Cancer Foundation supported the article processing fee of the journal. We thank all of the participants in the study. We also thank Formosa Medical Editors for help with manuscript preparation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Structure of 5 main parts in the first shared decision-making webpage.
[PNG File , 342 KB - jmir_v23i3e24926_app1.png]

Multimedia Appendix 2
Step 1 of the SDM process.
[PNG File , 273 KB - jmir_v23i3e24926_app2.png]

Multimedia Appendix 3
Step 2 of the SDM process.
[PNG File , 330 KB - jmir_v23i3e24926_app3.png]

Multimedia Appendix 4
Step 3 of the SDM process.
[PNG File , 367 KB - jmir_v23i3e24926_app4.png]

Multimedia Appendix 5
Step 4 of the SDM process.
[PNG File , 245 KB - jmir_v23i3e24926_app5.png]

References

1. The statistics and trends in health and welfare. Ministry of Health and Welfare Taiwan. 2018. URL: http://dep.mohw.gov.tw/DOS/np-1714-113.html [accessed 2020-02-24]

2. Fidler MM, Gupta S, Soerjomataram I, Ferlay J, Steliarova-Foucher E, Bray F. Cancer incidence and mortality among young adults aged 20–39 years worldwide in 2012: a population-based study. Lancet Oncol 2017 Dec;18(12):1579-1589. [doi: 10.1016/s1470-2045(17)30677-0]
3. Muñoz E, Domingo J, De Castro G, Lorenzo I, García-Velasco JA, Bellver J, et al. Ovarian stimulation for oocyte vitrification does not modify disease-free survival and overall survival rates in patients with early breast cancer. Reprod Biomed Online 2019 Nov;39(5):860-867. [doi: 10.1016/j.rbmo.2019.07.003] [Medline: 31564650]
4. Campbell-Enns HJ, Woodgate RL. The psychosocial experiences of women with breast cancer across the lifespan: a systematic review. Psychooncology 2017 Nov;26(11):1711-1721. [doi: 10.1002/pon.4281] [Medline: 27648726]
5. Ruggeri M, Pagan E, Bagnardi V, Bianco N, Gallerani E, Buser K, et al. Fertility concerns, preservation strategies and quality of life in young women with breast cancer: baseline results from an ongoing prospective cohort study in selected European Centers. Breast 2019 Oct;47:85-92. [doi: 10.1016/j.breast.2019.07.001] [Medline: 31362134]
6. Howard-Anderson J, Ganz PA, Bower JE, Stanton AL. Quality of life, fertility concerns, and behavioral health outcomes in younger breast cancer survivors: a systematic review. J Natl Cancer Inst 2012 Mar 07;104(5):386-405. [doi: 10.1093/jnci/djr541] [Medline: 22271773]
7. Fertility preservation. National Cancer Institute, National Institutes of Health. URL: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/fertility-preservation [accessed 2020-05-24]
8. Martinez M, Rabadan S, Domingo J, Cobo A, Pellicer A, Garcia-Velasco JA. Obstetric outcome after oocyte vitrification and warming for fertility preservation in women with cancer. Reprod Biomed Online 2014 Dec;29(6):722-728. [doi: 10.1016/j.rbmo.2014.09.002] [Medline: 25444506]
9. Oktay K, Harvey BE, Loren AW. Fertility preservation in patients with cancer: ASCO clinical practice guideline update summary. J Oncol Pract 2018 Jun;14(6):381-385. [doi: 10.1200/jop.18.00160]
10. Srikanthan A, Amir E, Bedard P, Giuliani M, Hodgson D, Laframboise S, et al. Fertility preservation in post-pubescent female cancer patients: a practical guideline for clinicians. Mol Clin Oncol 2018 Jan;8(1):153-158 [FREE Full text] [doi: 10.3892/mco.2017.1486] [Medline: 29387409]
11. Lambertini M, Moore HC, Leonard RC, Loibl S, Munster P, Bruzzone M, et al. Gonadotropin-releasing hormone agonists during chemotherapy for preservation of ovarian function and fertility in premenopausal patients with early breast cancer: a systematic review and meta-analysis of individual patient–level data. J Clin Oncol 2018 Jul 01;36(19):1981-1990. [doi: 10.1200/jco.2018.78.0858]
12. Hsieh PL, Huang SM, Chien LY, Lee CF, Hsiung Y, Tai CJ. Risk-benefit perception of pregnancy among breast cancer survivors. Eur J Cancer Care (Engl) 2018 Mar;27(2):e12696. [doi: 10.1111/ecc.12696] [Medline: 28440587]
13. Mehrabi E, Hajian S, Simbar M, Hoshyari M, Zayeri F. The lived experience of Iranian women confronting breast cancer diagnosis. J Caring Sci 2016 Mar;5(1):43-55 [FREE Full text] [doi: 10.15171/jcs.2016.005] [Medline: 26989665]
14. Villar RR, Fernández SP, Garea CC, Pillado MTS, Barreiro VB, Martín CG. Quality of life and anxiety in women with breast cancer before and after treatment. Rev Lat Am Enfermagem 2017 Dec 21;25:e2958 [FREE Full text] [doi: 10.1590/1518-8345.2258.2958] [Medline: 29267541]
15. Tsaras K, Papathanasiou I, Mitsi D, Veneti A, Kelesi M, Zyga S, et al. Assessment of depression and anxiety in breast cancer patients: prevalence and associated factors. Asian Pac J Cancer Prev 2018 Jun 25;19(6):1661-1669 [FREE Full text] [doi: 10.22034/APJCP.2018.19.6.1661] [Medline: 29938451]
16. Bártolo A, Santos IM, Valério E, Monteiro S. Depression and health-related quality of life among young adult breast cancer patients: the mediating role of reproductive concerns. J Adolesc Young Adult Oncol 2020 Jun;9(3):431-435. [doi: 10.1089/jayao.2019.0144] [Medline: 32208042]
17. Ussher JM, Parton C, Perz J. Need for information, honesty and respect: patient perspectives on health care professionals communication about cancer and fertility. Reprod Health 2018 Jan 05;15(1):2 [FREE Full text] [doi: 10.1186/s12978-017-0441-z] [Medline: 29304873]
18. Huang SM, Tseng LM, Lai JC, Lien PJ, Chen PH. Infertility-related knowledge in childbearing-age women with breast cancer after chemotherapy. Int J Nurs Pract 2019 Oct;25(5):e12765. [doi: 10.1111/ijn.12765] [Medline: 31313445]
19. Elwyn G, Laitner S, Coulter A, Walker E, Watson P, Thomson R. Implementing shared decision making in the NHS. BMJ 2010 Oct 14;341:c5146. [doi: 10.1136/bmj.c5146] [Medline: 20947577]
20. Elwyn G, Frosch DL, Kobrin S. Implementing shared decision-making: consider all the consequences. Implement Sci 2016 Aug 08;11(1):114 [FREE Full text] [doi: 10.1186/s13012-016-0480-9] [Medline: 27502770]
21. Stiggelbout AM, Pieterse AH, De HJCJM. Shared decision making: concepts, evidence, and practice. Patient Educ Couns 2015 Oct;98(10):1172-1179. [doi: 10.1016/j.pec.2015.06.022] [Medline: 26215573]
22. Truglio-Londrigan M, Slyer JT. Shared decision-making for nursing practice: an integrative review. Open Nurs J 2018 Jan;12(1):1-14 [FREE Full text] [doi: 10.2174/1874434601812010001] [Medline: 29456779]
23. Chang HL, Li FS, Lin CF. Factors influencing implementation of shared medical decision making in patients with cancer. Patient Prefer Adherence 2019 Nov;13:1995-2005. [doi: 10.2147/ppa.s217561]
24. Wiener RS, Koppelman E, Bolton R, Lasser KE, Borrelli B, Au DH, et al. Patient and clinician perspectives on shared decision-making in early adopting lung cancer screening programs: a qualitative study. J Gen Intern Med 2018 Jul;33(7):1035-1042 [FREE Full text] [doi: 10.1007/s11606-018-4350-9] [Medline: 29468601]

25. Zdenkowski N, Butow P, Tesson S, Boyle F. A systematic review of decision aids for patients making a decision about treatment for early breast cancer. Breast 2016 Apr;26:31-45. [doi: 10.1016/j.breast.2015.12.007] [Medline: 27017240]
26. Kemertzis MA, Ranjithakumaran H, Hand M, Peate M, Gillam L, McCarthy M, et al. Fertility Preservation Toolkit: a clinician resource to assist clinical discussion and decision making in pediatric and adolescent oncology. J Pediatr Hematol Oncol 2018;40(3):e133-e139. [doi: 10.1097/mph.0000000000001103]
27. Speller B, Metcalfe K, Kennedy ED, Facey M, Greenblatt E, Scheer AS, et al. The "Begin Exploring Fertility Options, Risks and Expectations" (BEFORE) decision aid: development and alpha testing of a fertility tool for premenopausal breast cancer patients. BMC Med Inform Decis Mak 2019 Oct 28;19(1):203. [doi: 10.1186/s12911-019-0912-y]
28. Garvelink MM, ter Kuile MM, Fischer MJ, Louwé LA, Hilders CGJM, Kroep JR, et al. Development of a decision aid about fertility preservation for women with breast cancer in The Netherlands. J Psychosom Obstet Gynaecol 2013 Dec;34(4):170-178. [doi: 10.3109/0167482X.2013.851663] [Medline: 24188788]
29. Fertility. Livestrong. URL: https://www.livestrong.org/what-we-do/program/fertility [accessed 2019-10-03]
30. Cordeiro L, Soares CB. Action research in the healthcare field. JBI Database System Rev Implement Rep 2018 Apr;16(4):1003-1047. [doi: 10.11124/jbisrir-2016-003200]
31. Bradbury H, Lifvergren S. Action research healthcare: focus on patients, improve quality, drive down costs. Healthc Manage Forum 2016 Nov;29(6):269-274. [doi: 10.1177/0840470416658905] [Medline: 27770047]
32. de Wit M, Beurskens A, Piškur B, Stoffers E, Moser A. Preparing researchers for patient and public involvement in scientific research: development of a hands-on learning approach through action research. Health Expect 2018 Aug;21(4):752-763 [FREE Full text] [doi: 10.1111/hex.12671] [Medline: 29418053]
33. Lee TI, Sheu SJ, Chang HC, Hung YT, Tseng LM, Chou SS, et al. Developing a web-based comic for newly diagnosed women with breast cancer: an action research approach. J Med Internet Res 2019 Feb 04;21(2):e10716 [FREE Full text] [doi: 10.2196/10716] [Medline: 30714947]
34. Hung YT, Wu CF, Liang TH, Chou SS, Chen GL, Wu PN, et al. Developing a decision-aid website for breast cancer surgery: an action research approach. J Med Internet Res 2019 Feb;21(2):e10404. [doi: 10.2196/10404]
35. Bennett C, Graham ID, Kristjansson E, Kearing SA, Clay KF, O'Connor AM. Validation of a preparation for decision making scale. Patient Educ Couns 2010 Jan;78(1):130-133. [doi: 10.1016/j.pec.2009.05.012] [Medline: 19560303]
36. The SHARE approach. Agency for Healthcare Research and Quality. 2019. URL: https://www.ahrq.gov/health-literacy/curriculum-tools/shareddecisionmaking/index.html [accessed 2020-06-24]
37. Taiwan internet report. Taiwan Network Information Center. 2020. URL: https://twnic.tw/stat_n.php [accessed 2020-05-24]
38. Huang SM, Tseng LM, Lai JC, Tsai YF, Lien PJ, Chen PH. Impact of symptom and social support on fertility intention in reproductive-age women with breast cancer. Clin Nurs Res 2020 Jul;29(6):411-418. [doi: 10.1177/1054773818770814] [Medline: 29682988]
39. Shared decision-making tool for fertility preservation. Comprehensive Breast Health Center Taipei Veterans General Hospital. URL: http://vghtpe-cbhc.sdmcare.me/healthEducation [accessed 2020-06-06]
40. Wigfall LT, Tanner AH. Literacy and health-care engagement as predictors of shared decision-making among adult information seekers in the USA: a secondary data analysis of the Health Information National Trends Survey. J Cancer Educ 2018 Feb;33(1):67-73. [doi: 10.1007/s13187-016-1052-z] [Medline: 27251634]
41. Wang MJ, Hung LC, Lo YT. Glycemic control in type 2 diabetes: role of health literacy and shared decision-making. Patient Prefer Adherence 2019 May;13:871-879. [doi: 10.2147/ppa.s202110]
42. Chew-Graham CA. Focus on shared decision-making. Health Expect 2017 Dec;20(6):1181-1182 [FREE Full text] [doi: 10.1111/hex.12654] [Medline: 29145699]

Abbreviations

IT: information technology
PrepDM: Preparation for Decision-Making scale
SDM: shared decision making

Edited by R Kukafka; submitted 24.10.20; peer-reviewed by S Sabarguna, A Abubakar; comments to author 28.12.20; revised version received 08.01.21; accepted 04.02.21; published 17.03.21.

Please cite as:
Tseng LM, Lien PJ, Huang CY, Tsai YF, Chao TC, Huang SM
Developing a Web-Based Shared Decision-Making Tool for Fertility Preservation Among Reproductive-Age Women With Breast Cancer: An Action Research Approach
J Med Internet Res 2021;23(3):e24926
URL: https://www.jmir.org/2021/3/e24926
doi:10.2196/24926
PMID:33729164

Original Paper

In-Clinic Versus Web-Based Multidisciplinary Exercise-Based Rehabilitation for Treatment of Low Back Pain: Prospective Clinical Trial in an Integrated Practice Unit Model

Kamshad Raiszadeh[1], MD; Jonathan Tapicer[1], MS; Lissa Taitano[1], DPT; Jonathan Wu[1], PA; Bahar Shahidi[2], PhD, DPT

[1]Spinezone Medical Fitness, San Diego, CA, United States

[2]Department of Orthopaedic Surgery, University of California San Diego, San Diego, CA, United States

Corresponding Author:
Bahar Shahidi, PhD, DPT
Department of Orthopaedic Surgery
University of California San Diego
9500 Gilman Dr
MC0863
San Diego, CA
United States
Phone: 1 858 822 0439
Email: bshahidi@health.ucsd.edu

Abstract

Background: The recent onset of the COVID-19 pandemic has highlighted the need to reduce barriers to access physical therapy and associated care through the use of web-based programs and telehealth for those seeking treatment for low back pain (LBP). Despite this need, few studies have compared the effectiveness of clinic-based versus web-based or telehealth services.

Objective: This study aims to compare the clinical outcomes of clinic-based multidisciplinary therapy in an integrated practice unit (C-IPU) model with online integrated multidisciplinary therapy (O-IPU) in individuals undergoing conservative care for LBP.

Methods: A total of 1090 participants were prospectively recruited to participate in a clinical trial registry (NCT04081896) through the SpineZone rehabilitation IPU program. All participants provided informed consent. Participants were allocated to the C-IPU (N=988) or O-IPU (N=102) groups based on their personal preferences. The C-IPU program consisted of a high-intensity machine-based core muscle resistance training program, whereas the O-IPU program consisted of therapist-directed home core strengthening exercises through a web-based platform. Changes in LBP symptom severity (Numeric Pain Rating Scale), disability (Oswestry Disability Index), goal achievement (Patient-Specific Functional Scale), and frequency of opioid use were compared between the C-IPU and O-IPU groups using multivariate linear regression modeling adjusted for age, gender, treatment number, program duration, and baseline pain and disability.

Results: Approximately 93.03% (1014/1090) of the participants completed their recommended programs, with no group differences in dropout rates (P=.78). The C-IPU group showed greater pain relief (P<.001) and reductions in disability (P=.002) than the O-IPU group, whereas the O-IPU group reported greater improvements in goal achievement (P<.001). Both programs resulted in reduced opioid use frequency, with 19.0% (188/988) and 21.5% (22/102) of participants reporting cessation of opioid use for C-IPU and O-IPU programs, respectively, leaving only 5.59% (61/1090) of participants reporting opioid use at the end of their treatment.

Conclusions: Both in-clinic and web-based multidisciplinary programs are beneficial in reducing pain, disability, and opioid use and in improving goal achievement. The differences between these self-selected groups shed light on patient characteristics, which require further investigation and could help clinicians optimize these programs.

Trial Registration: ClinicalTrials.gov NCT04081896; https://clinicaltrials.gov/ct2/show/NCT04081896

(J Med Internet Res 2021;23(3):e22548) doi:10.2196/22548

KEYWORDS

low back pain; telehealth; online therapy; physical therapy; integrated practice unit; rehabilitation

Introduction

Background

Low back pain (LBP) is a leading cause of disability globally, affecting people of all ages [1-3]. Clinic-based physical therapy visits and other physical activity programs have shown value and are currently the standard of care [4-7]. Internet- and web-based therapies have been increasingly used to implement physical rehabilitation and other behavioral programs [8,9]. Importantly, these platforms have the potential for widespread dissemination at a relatively low cost and convenience for users. This has become even more relevant, as health care practitioners and patients are navigating challenges associated with the COVID-19 pandemic. Recent policy changes during the pandemic have reduced the barriers to telehealth access and have promoted the use of telehealth and web-based platforms for primary, specialty care and physical therapy [10].

There is an abundance of commercially available apps offering pain management and exercises (eg, Kaia, Physera, Hinge, Curable, etc) for the treatment of LBP. In parallel, a small number of studies have demonstrated that telehealth and web-based platforms can be used to successfully perform health evaluations in individuals with chronic LBP [11,12]. However, to date, little research exists on the outcomes of internet-based physical activity treatment programs. Of the data that do exist, there is modest evidence for improvement in general health care outcomes based on smartphone app use, and systematic reviews have found weak evidence for the beneficial effects of digital interventions in LBP management [9,13,14]. Similar outcomes have been reported in other populations, such as those with heart failure [15] and knee pain [13,16].

Despite prior literature suggesting some clinical efficacy using telehealth or web-based platforms in individuals with LBP in the United States, there is little information on how these platforms compare to similar in-clinic programs. Most rehabilitation programs are administered by physical therapists in a clinical setting, and care is often not coordinated with the medical team, such as an integrated practice unit (IPU) with the overarching goal of high-value health care [17]. In an IPU, care is provided by practitioners with different specialties centered around the patient's disease process. Multidisciplinary spine programs have been shown to be more effective than physical therapy alone [18]; however, current studies have limited generalizability because of problems with access to and interpretation of evidence [19] and recruitment methods leading to populations that do not match general practice.

Objectives

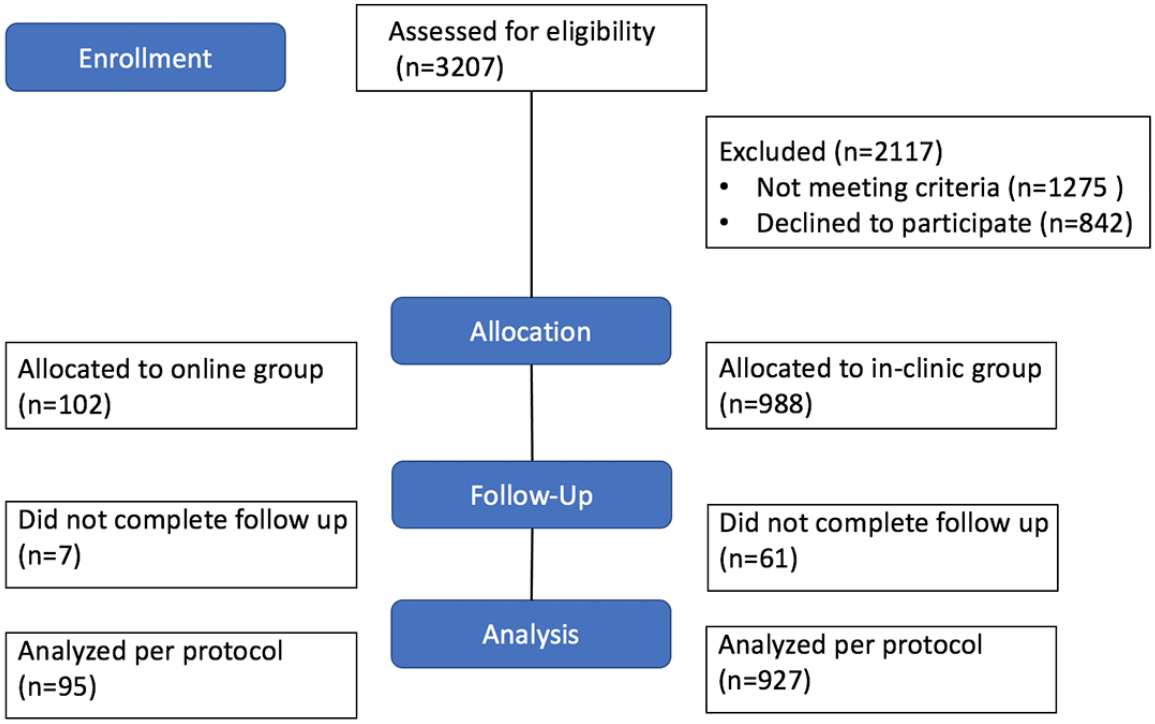
In this study, we compared the outcomes of in-clinic and web-based exercise-based multidisciplinary spinal treatment programs administered through an IPU. We hypothesized that in this model, both web-based and in-clinic treatment would result in equivalent improvement of patient outcomes of LBP-related symptom severity, disability, goal achievement, and frequency of opioid use. A secondary hypothesis was that individuals would self-allocate based on the severity of symptoms at baseline and that more complex or severely debilitated patients would be more likely to select the in-clinic program.

Methods

Study Population

This was a prospective cohort study using a consecutively enrolled convenience sample of individuals referred to the SpineZone rehabilitation program by their primary care physician. These participants were prospectively recruited to participate in a clinical trial registry (NCT04081896) between January 1 and June 30, 2019. All participants provided informed consent according to the approved institutional review board and the Declaration of Helsinki. Participants were eligible for inclusion if they were aged between 18 and 85 years and were seeking care for symptoms of LBP, including diagnoses of stenosis, disc degeneration, spondylolisthesis, scoliosis, vertebral fracture, radiculopathy, and nonspecific LBP. All participants who completed the initial evaluation questionnaires, including the Oswestry Disability Index (ODI) [20,21], the Numeric Pain Rating Scale (NPRS) [22,23] for pain, and a modified Patient-Specific Functional Scale (PSFS) [24] for goals before initiation of rehabilitation and participated in at least two sessions of either web-based or in-clinic treatment beyond the initial evaluation were included. The ODI is a self-report questionnaire that represents disability as a result of LBP and has been validated in this population [25-27]. The NPRS provides information on the intensity of pain experienced in the back or leg (in the case of radiating symptoms) and has also been validated in patients with LBP [22,23,28]. The PSFS is a self-report questionnaire that identifies patient-prioritized functional activities that are used to establish goals and has been validated in this population [24,29]. Participants self-allocated to either an in-clinic program or a web-based program based on personal preferences. Participants were educated about the time necessary to build muscle and encouraged to participate in as much of the 12-week program as possible. Therefore, to best evaluate the influence of a consistently administered program and to reduce the confounding effects of potential gaps in care on outcomes of interest, participants who had completed their treatment program within 6 months of initiation were included in the analysis (Figure 1). The duration of the program was allowed to vary according to the patient's needs, and the total number of visits and program duration was documented to account for this variability. For participants who reported symptom resolution before the recommended 12 weeks, the postassessment was conducted at the last attended visit.

Figure 1. Schematic of participant enrollment.



IPU Model

IPUs are a part of the strategy for high-value health care [17], along with measuring outcomes, bundled payments for the full cycle of care, integrated care, and ability for geographic outreach. Our IPU model consisted of a multidisciplinary treatment team, including physical therapists, orthopedic spine surgeons, spine-trained physician assistants, pain specialist consultants, and the clinical director for both in-clinic and web-based practitioners. Multidisciplinary conferences were carried out weekly with the team, and suggestions regarding different areas of focus, modification of exercise program or educational content, further evaluation, need for injections or medication, or need for surgical consultation were discussed and communicated with the primary physician. Physical therapy was administered by therapists trained in psychologically informed treatment strategies [30]. All available radiographic studies were reviewed by a physician assistant and surgeons, and patients were provided information on the natural history of common radiographic findings in the spine, as well as the prevalence of asymptomatic radiologic findings.

Patients in both programs were provided with customized educational materials regarding their condition, sleep, mindfulness, nutrition, posture, and ergonomics via the web-based application. Psychosocial risk factors include catastrophizing, fear avoidance, magnification, depression, and anxiety [30]. At-risk patients underwent education on expressive writing and cognitive behavioral techniques [31]. Group mindfulness classes were also encouraged in patients who were thought to exhibit psychosocial risk factors.

Clinic-Based IPU Group Protocol

Participants allocated to the in-clinic IPU program (C-IPU) underwent an initial history and standard physical examination by a licensed physical therapist. This examination also included a postural assessment and measurement of isometric lumbar extension strength using a Med-X isokinetic dynamometer [32]. On the basis of these initial assessments, the physical therapist prescribed and progressed machine-based resistance core strengthening exercises, as previously described in detail [33].

Online IPU Group Protocol

Participants allocated to the online IPU (O-IPU) underwent an initial history and virtual physical examination by a licensed physical therapist or physical therapy aide. On the basis of general goals (activities of daily living improvement or sports and performance), symptom presentation (back pain, back pain with radiating symptoms, and back pain with radiating symptoms that improved with extension), and acuity of symptoms (acute versus chronic), patients were assigned specific directional preference exercises and core strengthening exercises that best matched their symptomatology. Directional preference and home core strengthening exercises were administered on a customized web-based platform along with condition-specific educational content. The exercises assigned on the web-based platform were implemented using a custom mobile-enabled platform that provided images and videos of exercises with written instructions and tracking for a number of sets and repetitions. The number of times logged in and whether the patient accessed the educational material was documented through the platform.

Outcome Measures and Statistical Analysis

The primary predictor variable of interest was group assignment (web-based vs in-clinic). The primary outcomes of interest included improvements in pain as measured by the change in NPRS score, improvements in LBP-related disability as measured by change in ODI, improvements in goal achievement as measured by change in the PSFS score, and changes in pain medication usage as measured by the frequency of opioid usage according to the following categories: none, <1 per day, 1 to 2 per day, 3 to 5 per day, and >6 per day. Demographic variables such as age, gender, chronicity of symptoms, and history of prior treatment were collected. Chronicity of symptoms was dichotomized into acute (≤3 months duration) and chronic (>3 months duration) symptoms, and prior treatments were categorized as none, conservative care (1, 2, or 3 different modalities), injections, or spinal surgery. Conservative care modalities were defined as physical therapy, acupuncture, or chiropractic treatment. All information was collected by the software platform and confirmed by the physical therapist performing the initial evaluations, re-evaluations, and discharge assessments. For the continuous descriptive variables (demographics and baseline characteristics) and outcomes of interest (change in pain, disability, and goal achievement), univariate and adjusted linear regression models were generated for each outcome, with adjustments including covariates of age, number of visits, total program duration, and baseline levels of pain, disability, or goals. For the categorical or binary descriptive variables (demographics and baseline characteristics) and outcomes of interest (opioid use frequency), chi-square or logistic regression models were used on an as-treated basis. All statistical analyses were performed using SPSS software (version 26.0.0). Statistical significance was set to a *P* value of .05.

Results

Participant Characteristics

A total of 1090 participants were included in the analysis. The average age of participants was 62.3 (SD 16.49) years, with 58.81% (641/1090) of participants being female. The mean pain levels were moderate (4.96, SD 1.78 points), as was the average LBP-related disability (26.92, SD 14.26 points) at baseline. Most participants (948/1090, 86.97%) reported symptom durations of greater than 3 months across both groups, indicating chronic symptoms. There were no differences in the proportion of chronic symptoms between the groups (*P*=.47). Most participants reported seeking other treatments before initiating the current program, with 30.46% (332/1090) having received a single modality of prior conservative treatment (physical therapy, chiropractic, or acupuncture), 10.55% (115/1090) having received 2 modalities, and 5.41% (59/1090) having received 3 or more modalities. Some participants reported more invasive prior interventions, such as injections (190/1090, 17.43%) or prior spine surgery (77/1090, 7.06%). Approximately one-third (317/1090, 29.08%) of the participants reported no prior treatment. A greater proportion of participants in the web-based group had received prior treatment (*P*<.001), with more having received 2 (*P*<.001) or ≥3 (*P*=.03) different modalities of conservative treatment. There were no differences in the proportions of participants who had received a single conservative treatment modality (*P*=.85), injections (*P*=.42), or surgery (*P*=.09) between the groups. In addition, the majority (821/1090, 75.32%) of participants were not taking any opioid medications at the time of initiating treatment, with 11.46% (125/1090) reporting taking opioids <1 per day, 7.98% (87/1090) taking 1 to 2 per day, 4.22% (46/1090) taking 3 to 5 per day, and 1.00% (11/1090) taking ≥6 per day. There were no significant differences in opioid use frequency at baseline between groups (*P*=.29). The in-clinic group reported higher levels of pain (*P*=.03) and LBP-related disability (*P*<.001) at baseline than the web-based group (but had higher scores on their goal achievement at baseline (*P*<.001). The participants who enrolled in the web-based program participated in the program for a shorter duration (*P*<.001) but participated in more visits than the in-clinic participants (*P*<.001). More participants in the O-IPU group were diagnosed with lumbar radiculopathy, and more participants in the C-IPU group were diagnosed with degenerative disc disease (*P*<.001). A comparison of the baseline characteristics for each group is shown in Table 1.

Table 1. Baseline characteristics between online integrated practice unit and in-clinic integrated practice unit groups.

Variable and group	Value	*t* test or χ^2 (*df*)	*P* value	Difference
Numeric Pain Rating Scale initial (points), mean (SD)		−2.25 (1088)	*.03* [a]	−0.42
Web-based	4.58 (1.73)			
In clinic	5.00 (1.78)			
Oswestry Disability Index initial (points), mean (SD)		−3.99 (1088)	*<.001* [a]	−5.87
Web-based	21.50 (12.09)			
In clinic	27.47 (14.36)			
Patient-Specific Functional Scale score (points), mean (SD)		−3.74 (1088)	*<.001* [a]	−0.80
Web-based	3.10 (2.11)			
In clinic	3.90 (2.04)			
Visit number (days or log-ins), mean (SD)		5.92 (1088)	*<.001* [a]	12.56
Web-based	36.62 (9.87)			
In clinic	24.06 (21.18)			
Program duration (days), mean (SD)		−13.55 (1088)	*<.001* [a]	−46.47
Web-based	45.73 (6.49)			
In clinic	92.19 (34.55)			
Age (years), mean (SD)		1.38 (1088)	.17	2.37
Web-based	64.41 (11.79)			
In clinic	62.04 (16.89)			
Gender (female), n (%)		2.18 (1)	.14	8.31
Web-based, n=102	52 (51.20)			
In clinic, n=988	588 (59.50)			
Symptom duration >3 months, n (%)		0.50 (1)	.50	2.50
Web-based, n=102	91 (89.20)			
In clinic, n=988	857 (86.70)			
Opioid use frequency initial, n (%)				
None		0.001 (1)	.97	−0.19
Web-based, n=102	77 (75.49)			
In clinic, n=988	744 (75.30)			
<1 per day		0.05 (1)	.82	0.72
Web-based, n=102	11 (10.78)			
In clinic, n=988	114 (11.50)			
1 to 2 per day		0.68 (1)	.41	2.32
Web-based, n=102	6 (5.88)			
In clinic, n=988	81 (8.20)			
3 to 5 per day		0.13 (1)	.72	−0.75
Web-based, n=102	5 (4.90)			
In clinic, n=988	41 (4.15)			
>6 per day		2.34 (1)	.13	−2.13
Web-based, n=102	3 (2.94)			
In clinic, n=988	8 (0.81)			
Prior treatments, n (%)				

Variable and group	Value	*t* test or χ^2 (*df*)	*P* value	Difference
None		78.82 (1)	*<.001* [a]	19.10
Web-based, n=102	12 (11.80)			
In clinic, n=988	305(30.90)			
1 modality		0.04 (1)	.85	2.30
Web-based, n=102	29(28.40)			
In clinic, n=988	303(30.70)			
2 modalities		10.93 (1)	*<.001* [a]	−10.00
Web-based, n=102	20(19.60)			
In clinic, n=988	95(9.60)			
3 or more modalities		4.75 (1)	*.03* [a]	−4.80
Web-based, n=102	10(9.80)			
In clinic, n=988	49(5.00)			
Injections		0.64 (1)	.42	−2.40
Web-based, n=102	20(19.60)			
In clinic, n=988	170(17.20)			
Surgery		2.84 (1)	.09	−4.10
Web-based, n=102	11(10.80)			
In clinic, n=988	66(6.70)			
Diagnosis, n (%)				
Lumbar radiculopathy		40.51 (1)	*<.001* [a]	−21.00
Web-based, n=102	32(31.40)			
In clinic, n=988	103 (10.40)			
Nonspecific low back pain		1.70 (1)	.19	−20.10
Web-based, n=102	54 (52.90)			
In clinic, n=988	324 (32.80)			
Spondylolisthesis		3.34 (1)	.07	−5.80
Web-based, n=102	4 (3.90)			
In clinic, n=988	96 (9.70)			
Stenosis		3.10 (1)	.08	6.30
Web-based, n=988	96 (5.90)			
In clinic, n=988	119 (12.20)			
Scoliosis		2.69 (1)	.10	3.40
Web-based, n=102	1 (1.00)			
In clinic, n=988	43 (4.40)			
Degenerative disc disease		25.32 (1)	*<.001* [a]	23.70
Web-based, n=102	5 (4.90)			
In clinic, n=988	283 (28.60)			
Fracture		0.40 (1)	.40	1.90
Web-based, n=102	0 (0.00)			
In clinic, n=988	19 (1.90)			

[a]Italics indicate significant difference between online and in clinic groups.

Clinical Outcomes

More than 93% of participants completed their recommended program, with no differences in dropout between the groups (*P*=.78). For the primary outcome of pain improvement, both groups achieved clinically significant reductions, with the in-clinic group demonstrating a significantly greater improvement in pain compared with the web-based group. However, although these differences were statistically significant, they were not clinically significant (mean difference 1.02 points, SE 1.36; *P*<.001; Figure 2). Similarly, the in-clinic group demonstrated statistically larger improvements in LBP-related disability, but these group differences did not reach clinical significance (mean difference 4.26 points, SE 0.32; *P*=.002; Figure 2) .Overall, participants achieved 15.62% (SD 63.6) reductions in LBP-related disability from their baseline scores. Despite greater improvements in pain and disability in the C-IPU group, the O-IPU group reported greater improvements in goal achievement (mean difference 1.70; *P*<.001; Figure 3). The mean change scores and results for the univariate comparisons are reported in Tables 2 and 3. These findings did not change when the models were corrected for age, program visits or duration, or levels of baseline pain or disability, with the exception of goal achievement, where the group differences lost significance but retained a trend (*P*=.06; Table 4). Finally, we observed substantial reductions in the frequency of opioid use for both the in-clinic and web-based programs, with 18.92% (187/988) and 21.56% (22/102) of participants reporting cessation of opioid use on completion of the in-clinic and web-based programs, respectively, leaving only 5.59% (61/1090) of participants reporting opioid use at the end of their treatment (Figure 3). There were no significant differences in the reduction in opioid use frequency between the groups (*P*=.97).

Figure 2. Mean and SDs for low back pain (A) and disability (B) between O-IPU (online) and C-IPU (in-clinic) groups at treatment baseline and discharge. IPU: integrated practice unit; NPRS: numeric pain rating scale; ODI: Oswestry disability index.

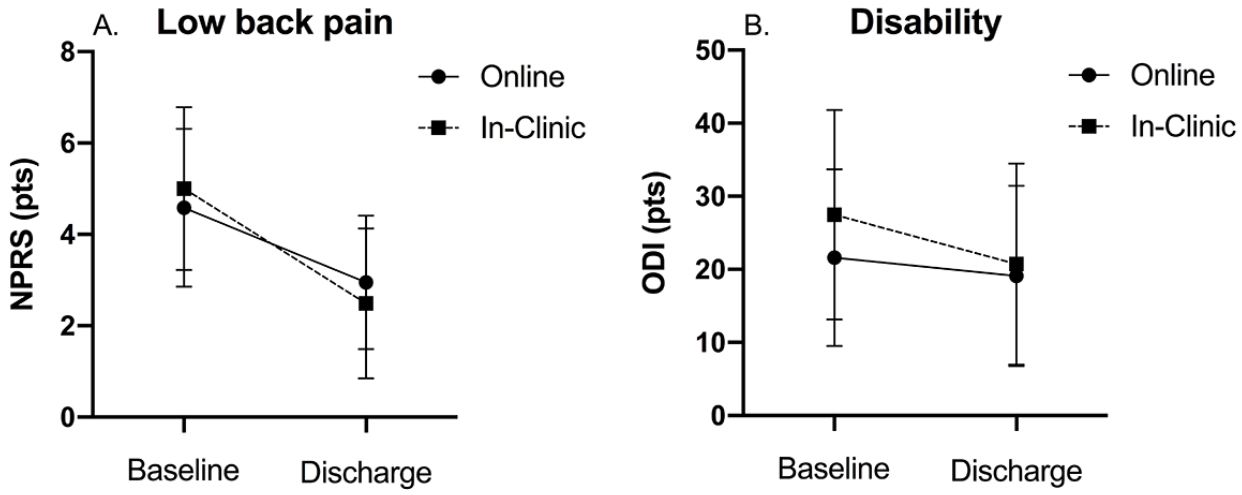


Figure 3. Means and standard deviations for patient opioid use frequency (A) and goal achievement (B) between O-IPU (online) and C-IPU (in-clinic) groups at treatment baseline and discharge. IPU: integrated practice unit; PSFS: patient-specific functional scale.

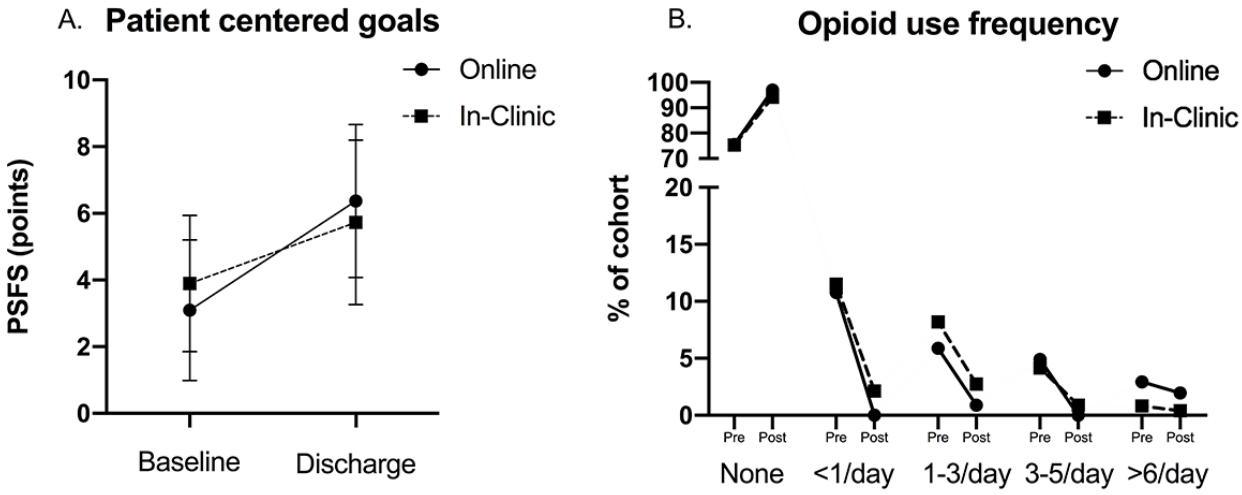


Table 2. Unadjusted comparisons for primary outcomes of interest between online integrated practice unit and in-clinic integrated practice unit groups.

Variable and group	Value, mean (SD)	*t* test (*df*)	*P* value	Difference
Change in pain (points)		4.64 (1020)	*<.001* [a]	1.02
Web-based	−1.97 (1.57)			
In clinic	−2.99 (2.10)			
Change in Oswestry Disability Index (points)		3.15 (1020)	*.002* [a]	4.26
Web-based	−2.97 (10.96)			
In clinic	−7.23 (12.93)			
Change in Patient-Specific Functional Scale (points)		5.38 (1020)	*<.001* [a]	1.70
Web-based	3.14 (3.13)			
In clinic	1.44 (2.72)			

[a]Italics indicate significant difference between online and in clinic groups.

Table 3. Unadjusted comparisons for change in opioid use frequency between online integrated practice unit and in-clinic integrated practice unit groups.

Variable	None	1 to 3 per day	3 to 5 per day	>6 per day	χ^2*(df)*	*P* value
Change in opioid use frequency, %					0.56 (3)	.97
Web-based	21.51	−4.98	−4.90	−0.98		
In clinic	18.83	−5.47	−3.25	−0.41		

Table 4. Results of the multivariate linear regression for primary outcomes of interest.

Model	ß coefficient	SE	*t* test (*df*)	*P* value
Pain improvement				
(Constant)	.49	0.31	1.60 (5)	.11
In clinic	−1.04	0.19	−5.36 (5)	*<.001* [a]
Program duration (days)	.01	0.002	3.32 (5)	*.001* [a]
Visit number (days per log-ins)	0	0.002	0.16 (5)	.87
Age (years)	.01	0.003	2.64 (5)	*.009* [a]
Baseline pain (points)	−.67	0.03	−24.00 (5)	*<.001* [a]
Oswestry Disability Index improvement				
(Constant)	2.74	1.92	1.43 (5)	.15
In clinic	−3.01	1.31	−2.30 (5)	*.02* [a]
Program duration (days)	.03	0.01	2.40 (5)	*.02* [a]
Visit number (days/log-ins)	−.03	0.02	−1.48 (5)	.14
Age (years)	.07	0.02	3.02 (5)	*.003* [a]
Baseline Oswestry Disability Index (points)	−.46	0.03	−18.17 (5)	*<.001* [a]
Goals improvement				
(Constant)	6.40	0.43	14.82 (5)	*<.001* [a]
In clinic	−.52	0.28	−1.86 (5)	.06
Program duration (days)	−.01	0.002	−2.12 (5)	*.03* [a]
Visit number (days per log-ins)	.01	0.004	1.54 (5)	.13
Age (years)	−.01	0.005	−2.50 (5)	*.01* [a]
Baseline goals (points)	−.77	0.04	−20.97 (5)	*<.001* [a]

[a]Italics indicate significant difference between online and in clinic groups.

Discussion

Principal Findings

This study demonstrates that both the in-clinic and web-based multidisciplinary programs administered by an IPU resulted in reductions in pain, LBP-related disability, and opioid use in individuals seeking conservative management for LBP. However, contrary to our hypothesis, the in-clinic program demonstrated statistical superiority over the web-based program for pain and disability, although these differences did not reach clinical significance. Similarly, although both groups demonstrated improvements in patient-centered goals, the web-based group reported larger improvements in goal achievement. Importantly, this is the first study to demonstrate reductions in opioid usage in addition to symptom- and function-based outcome measures using a web-based platform.

Comparisons Between the C-IPU and O-IPU Populations

The results of this study confirm our secondary hypothesis that individuals who self-allocate to the in-clinic program have more severe symptomatology at baseline than those who allocate to the web-based program, with the in-clinic participants demonstrating higher levels of baseline pain and disability. This may explain the larger improvements in pain and disability in the C-IPU group, although the retention of group differences after correcting for baseline pain and disability in the multivariate model suggests that this may not be the only explanation. Similarly, the observation that a greater proportion of participants in the O-IPU group had received multiple different modalities of prior conservative treatment compared with the C-IPU group suggests that this population has an interest in exploring alternative treatments to achieve highly personalized goals. The finding that the O-IPU group reported larger improvements in goal achievement supports the concept that individuals who self-select web-based platforms may have different goals and expectations at baseline than those coming

to the clinic. In addition, the underlying bias or belief in treatment success of one treatment modality over the other may influence patient-reported outcomes. Further studies are needed to better explore these patient-selected preferences.

Comparisons Between Clinical Outcomes or Effectiveness and Prior Studies

Few prior studies have compared the effectiveness of web-based and in-clinic rehabilitation in individuals with LBP, and even fewer studies have incorporated a true multidisciplinary component in their program. One prospective single-arm study investigating program compliance and improvements in LBP and knee pain with a 12-week multidisciplinary digital care program incorporating education and sensor-guided exercise therapy (Hinge) and behavioral support with one-on-one remote health coaching found a significant relationship between app engagement and pain reduction [13]. However, no functional outcomes were obtained, and the population was recruited from employees through email, direct mail, and posters, which may not necessarily represent the general clinical population seeking treatment for LBP.

Of the studies comparing web-based and standard treatments, the results are conflicting. Toelle et al [14] performed a randomized controlled trial comparing an app-based (Kaia) back pain treatment program with a combination of physiotherapy and web-based education and found that the app-based treatment resulted in greater improvements in pain but no group differences in functional ability. However, the treatment frequency in the app-based group was 3 times per week for 12 weeks, whereas the treatment frequency in the physiotherapy group was 1 visit per week for 6 weeks, which may have resulted in an exposure bias in favor of the web-based platform. Indeed, when pain was compared at the 6-week time point (at the end of the physiotherapy group treatment program), both groups demonstrated similar symptoms. Other studies demonstrated no differences across treatment groups; in a randomized controlled trial comparing a web-based app (FitBack) with a wait-listed control group and an alternative care group receiving web-based educational materials via email [34], although the app-based treatment resulted in significantly lower odds of reporting back pain, along with improved functionality, quality of life, and well-being at 4 months posttreatment compared with the control group, there were no differences in these outcomes compared with the alternative care group. Similarly, Mbada et al [35] compared clinic-based McKenzie therapy versus telerehabilitation and found no significant difference in pain, disability, or quality of life between treatment groups. Of note, the McKenzie-based directional preference exercises were also used in this study for patients suspected of having disc pathology irritating or compressing neural structures.

Patient Population and Treatment Program Methodology

The results of this study demonstrating that the in-clinic program demonstrates statistical superiority for the outcomes of pain and disability are in contrast to other studies reporting the equivalence or superiority of a web-based program. These differences may be because of the patient populations recruited as well as the program design and comparison groups studied. For example, all patients in this study were referred to the program by their primary care physician after the failure of initial treatment with anti-inflammatory medication and education. In prior studies, participants were recruited using methods such as Facebook or other web-based advertisements [14], employer referrals [13], or employer wellness programs [34]. These recruitment methods may not be as generalizable to the standard population of patients with LBP seen by primary care physicians in medical group settings. In addition, one study excluded participants who had received medical care before enrolling in an intervention [34]. Overall, these recruitment methods may have resulted in a selection bias toward a more acute or less severe patient population. Indeed, Toelle et al [14] acknowledged that their study population demonstrated high levels of functional ability in both groups at baseline.

Another difference between this study and prior literature is the program design and methodology. Although many of the app-based platforms incorporate various factors related to back pain within the context of a biopsychosocial disease model, the use of a multidisciplinary approach for exercise-based rehabilitation with continued feedback through active engagement of an integrated care team has not been investigated in prior literature. For example, Bailey et al [13] used a sensor-guided exercise program as well as one-on-one remote health coaching using certified health coaches (through National Board for Health and Wellness Coaching), but patients were not continuously monitored by a multidisciplinary team over the course of their treatment. Similarly, Irvine et al [34] and Toelle et al [14] used predominantly app-based treatment and physical therapists for the control group but did not incorporate routine monitoring by other care providers as part of the treatment progress. In this study, both groups underwent physical therapy administered by psychologically informed practitioners [30], and patients who did not progress were reviewed in a weekly multidisciplinary conference with the physician assistants and surgeons to make adjustments to care, including the need for diagnostic studies, injections, or surgical intervention.

Study Limitations

This study had several limitations. First, it did not employ a true no-treatment control group, making the natural history effects of the treatment difficult to rule out. However, given that the goal of this study was to determine whether web-based implementation would provide similar benefits to in-clinic rehabilitation, the lack of a control group should not influence the primary study hypothesis. Second, this study employed a pragmatic study design, in which participants were not randomly allocated to treatment groups, introducing the possibility of selection bias. Indeed, some differences in baseline characteristics (eg, pain, disability) were observed between the groups. However, our statistical approach of adjusting for these baseline differences allowed us to correct for some of these discrepancies. Second, it also allows us to gain a better understanding of the factors influencing patient preferences in choosing care. Finally, although both groups experienced reduced pain, opioid use, and improved goal achievement that reached clinical relevance (determined by minimal clinically important difference values) [20-25,28], the reductions in

disability did not reach clinical significance. However, given the concurrent reductions in opioid use in a proportion of patients, the overall reductions in pain and disability may be underestimated because of decreases in pharmacological management.

Conclusions

This study found that C-IPU and O-IPU programs administered by a multidisciplinary team in an IPU both resulted in reductions in symptom severity, LBP-related disability, and opioid use frequency as well as improvements in goal achievement. The C-IPU was statistically superior to the O-IPU group in reducing pain and disability, and the O-IPU group was statistically superior in improving patient-specific goal achievement. Both programs resulted in equivalent and substantial reduction in opioid use frequency, which is a priority area in a population that is at high risk for developing opioid dependence.

Acknowledgments

This study was supported by grant funding from the Foundation for Physical Therapy Research, Magistro Family Foundation, awarded to BS.

Conflicts of Interest

KR owns stocks and serves as Chief Medical Officer for SpineZone Medical Fitness, and JT owns stocks and serves as Chief Technology Officer for SpineZone Medical Fitness. BS is a consultant for the San Diego Spine Foundation.

References

1. Andersson GB. Epidemiological features of chronic low-back pain. Lancet 1999 Aug 14;354(9178):581-585. [doi: 10.1016/S0140-6736(99)01312-4] [Medline: 10470716]
2. Lancet: Department of Error. 2017. URL: https://www.thelancet.com/journals/lancet/issue/vol390no10106/PIIS0140-6736(17)X0047-0 [accessed 2021-02-13]
3. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. Lancet 2017;389(10070):736-747. [doi: 10.1016/s0140-6736(16)30970-9]
4. Chou R. Low back pain (chronic). BMJ Clin Evid 2010 Oct 08;2010:- [FREE Full text] [Medline: 21418678]
5. Chou R, Huffman LH, American Pain Society, American College of Physicians. Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. Ann Intern Med 2007 Oct 02;147(7):492-504 [FREE Full text] [doi: 10.7326/0003-4819-147-7-200710020-00007] [Medline: 17909210]
6. Chou R, Qaseem A, Snow V, Casey D, Cross JT, Shekelle P, Clinical Efficacy Assessment Subcommittee of the American College of Physicians, American College of Physicians, American Pain Society Low Back Pain Guidelines Panel. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. Ann Intern Med 2007 Oct 02;147(7):478-491 [FREE Full text] [doi: 10.7326/0003-4819-147-7-200710020-00006] [Medline: 17909209]
7. Foster NE, Anema JR, Cherkin D, Chou R, Cohen SP, Gross DP, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet 2018 Jun;391(10137):2368-2383. [doi: 10.1016/s0140-6736(18)30489-6]
8. Physicians' motivations and requirements for adopting digital health Adoption and attitudinal shifts from 2016 to 2019. AMA Digital Health Research. 2020. URL: https://www.ama-assn.org/system/files/2020-02/ama-digital-health-study.pdf [accessed 2021-02-13]
9. Garg S, Garg D, Turin TC, Chowdhury MFU. Web-based interventions for chronic back pain: a systematic review. J Med Internet Res 2016 Jul 26;18(7):e139 [FREE Full text] [doi: 10.2196/jmir.4932] [Medline: 27460413]
10. Bashshur R, Doarn CR, Frenk JM, Kvedar JC, Woolliscroft JO. Telemedicine and the covid-19 pandemic, lessons for the future. Telemed J E Health 2020 May 01;26(5):571-573. [doi: 10.1089/tmj.2020.29040.rb] [Medline: 32275485]
11. Palacín-Marín F, Esteban-Moreno B, Olea N, Herrera-Viedma E, Arroyo-Morales M. Agreement between telerehabilitation and face-to-face clinical outcome assessments for low back pain in primary care. Spine (Phila Pa 1976) 2013 May 15;38(11):947-952. [doi: 10.1097/BRS.0b013e318281a36c] [Medline: 23238489]
12. Truter P, Russell T, Fary R. The validity of physical therapy assessment of low back pain via telerehabilitation in a clinical setting. Telemed J E Health 2014 Feb;20(2):161-167. [doi: 10.1089/tmj.2013.0088] [Medline: 24283249]
13. Bailey JF, Agarwal V, Zheng P, Smuck M, Fredericson M, Kennedy DJ, et al. Digital care for chronic musculoskeletal pain: 10,000 participant longitudinal cohort study. J Med Internet Res 2020 May 11;22(5):e18250 [FREE Full text] [doi: 10.2196/18250] [Medline: 32208358]
14. Toelle TR, Utpadel-Fischler DA, Haas K, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019 May 3;2(1):34 [FREE Full text] [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]

15. O'Connor M, Asdornwised U, Dempsey M, Huffenberger A, Jost S, Flynn D, et al. Using telehealth to reduce all-cause 30-day hospital readmissions among heart failure patients receiving skilled home health services. Appl Clin Inform 2017 Dec 16;07(02):238-247. [doi: 10.4338/aci-2015-11-soa-0157]
16. Manske RC, Lehecka BJ. Evidence - based medicine/practice in sports physical therapy. Int J Sports Phys Ther 2012 Oct;7(5):461-473 [FREE Full text] [Medline: 23091778]
17. Porter ME, Lee TH. The strategy that will fix health care. Harvard Business Review. 2013. URL: https://hbr.org/2013/10/the-strategy-that-will-fix-health-care [accessed 2021-02-13]
18. Davin S. Comparative effectiveness of an interdisciplinary pain program for chronic low back pain, compared to physical therapy alone. Spine (Phila Pa 1976) 2019;44(24):1715-1722. [doi: 10.1097/brs.0000000000003161]
19. Maher CG, Sherrington C, Elkins M, Herbert RD, Moseley AM. Challenges for evidence-based physical therapy: accessing and interpreting high-quality evidence on therapy. Phys Ther 2004 Jul;84(7):644-654. [Medline: 15225083]
20. Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. Phys Ther 2002 Jan;82(1):8-24. [doi: 10.1093/ptj/82.1.8] [Medline: 11784274]
21. Chiarotto A, Maxwell LJ, Terwee CB, Wells GA, Tugwell PJ, Ostelo RW. Roland-morris disability questionnaire and oswestry disability index: which has better measurement properties for measuring physical functioning in nonspecific low back pain? Systematic review and meta-analysis. Phys Ther 2016 Oct;96(10):1620-1637. [doi: 10.2522/ptj.20150420] [Medline: 27081203]
22. Chiarotto A, Maxwell LJ, Ostelo RW, Boers M, Tugwell P, Terwee CB. Measurement properties of visual analogue scale, numeric rating scale, and pain severity subscale of the brief pain inventory in patients with low back pain: a systematic review. J Pain 2019 Mar;20(3):245-263. [doi: 10.1016/j.jpain.2018.07.009] [Medline: 30099210]
23. Childs JD, Piva SR, Fritz JM. Responsiveness of the numeric pain rating scale in patients with low back pain. Spine (Phila Pa 1976) 2005 Jun 01;30(11):1331-1334. [doi: 10.1097/01.brs.0000164099.92112.29] [Medline: 15928561]
24. Horn KK, Jennings S, Richardson G, Vliet DV, Hefford C, Abbott JH. The patient-specific functional scale: psychometrics, clinimetrics, and application as a clinical outcome measure. J Orthop Sports Phys Ther 2012 Jan;42(1):30-42. [doi: 10.2519/jospt.2012.3727] [Medline: 22031594]
25. Fairbank JCT, Pynsent PB. The Oswestry Disability Index. Spine (Phila Pa 1976) 2000 Nov 15;25(22):2940-2952. [doi: 10.1097/00007632-200011150-00017] [Medline: 11074683]
26. Johnsen LG, Hellum C, Nygaard OP, Storheim K, Brox JI, Rossvoll I, et al. Comparison of the SF6D, the EQ5D, and the Oswestry Disability Index in patients with chronic low back pain and degenerative disc disease. BMC Musculoskelet Disord. 2013. URL: https://www.medscape.com/viewarticle/804040 [accessed 2021-02-13]
27. Vianin M. Psychometric properties and clinical usefulness of the Oswestry Disability Index. J Chiropr Med 2008 Dec;7(4):161-163 [FREE Full text] [doi: 10.1016/j.jcm.2008.07.001] [Medline: 19646379]
28. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken) 2011 Nov 07;63 Suppl 11(S11):S240-S252 [FREE Full text] [doi: 10.1002/acr.20543] [Medline: 22588748]
29. Maughan EF, Lewis JS. Outcome measures in chronic low back pain. Eur Spine J 2010 Sep 17;19(9):1484-1494 [FREE Full text] [doi: 10.1007/s00586-010-1353-6] [Medline: 20397032]
30. Psychologically informed physical therapy: pragmatic application for low back pain. University of Pittsburgh. 2016. URL: https://docplayer.net/77351746-Learning-objectives-psychologically-informed-physical-therapy-pipt-pragmatic-application-for-low-back-pain-3-7-2016.html [accessed 2021-02-13]
31. Pepe L, Milani R, Di Trani M, Di Folco G, Lanna V, Solano L. A more global approach to musculoskeletal pain: expressive writing as an effective adjunct to physiotherapy. Psychol Health Med 2014 Nov 20;19(6):687-697. [doi: 10.1080/13548506.2013.859712] [Medline: 24251961]
32. Nelson BW, Carpenter DM, Dreisinger TE, Mitchell M, Kelly CE, Wegner JA. Can spinal surgery be prevented by aggressive strengthening exercises? A prospective study of cervical and lumbar patients. Arch Phys Med Rehabil 1999 Jan;80(1):20-25. [doi: 10.1016/s0003-9993(99)90302-7] [Medline: 9915367]
33. Berry DB, Padwal J, Johnson S, Englund EK, Ward SR, Shahidi B. The effect of high-intensity resistance exercise on lumbar musculature in patients with low back pain: a preliminary study. BMC Musculoskelet Disord 2019 Jun 18;20(1):290 [FREE Full text] [doi: 10.1186/s12891-019-2658-1] [Medline: 31208400]
34. Irvine AB, Russell H, Manocchia M, Mino DE, Cox Glassen T, Morgan R, et al. Mobile-Web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015 Jan 02;17(1):e1 [FREE Full text] [doi: 10.2196/jmir.3130] [Medline: 25565416]
35. Mbada CE, Olaoye MI, Dada OO, Ayanniyi O, Johnson OE, Odole AC, et al. Comparative efficacy of clinic-based and telerehabilitation application of Mckenzie therapy in chronic low-back pain. Int J Telerehabil 2019;11(1):41-58 [FREE Full text] [doi: 10.5195/ijt.2019.6260] [Medline: 31341546]

Abbreviations

C-IPU: clinic integrated practice unit
IPU: integrated practice unit
LBP: low back pain
NPRS: Numeric Pain Rating Scale
ODI: Oswestry Disability Index
O-IPU: online integrated practice unit
PSFS: Patient-Specific Functional Scale

Edited by R Kukafka; submitted 15.07.20; peer-reviewed by D Lipardo, F Pereira; comments to author 14.09.20; revised version received 02.10.20; accepted 31.01.21; published 18.03.21.

Please cite as:
Raiszadeh K, Tapicer J, Taitano L, Wu J, Shahidi B
In-Clinic Versus Web-Based Multidisciplinary Exercise-Based Rehabilitation for Treatment of Low Back Pain: Prospective Clinical Trial in an Integrated Practice Unit Model
J Med Internet Res 2021;23(3):e22548
URL: https://www.jmir.org/2021/3/e22548
doi:10.2196/22548
PMID:33734088

Original Paper

Attitudes and Engagement of Pregnant and Postnatal Women With a Web-Based Emotional Health Tool (Mummatters): Cross-sectional Study

Nicole Reilly[1,2*], BPsyc, PhD; Marie-Paule Austin[2,3*], MBBS, MD

[1]Centre for Health Service Development, Australian Health Services Research Institute, University of Wollongong, Wollongong, Australia

[2]Perinatal & Women's Mental Health Unit, St John of God Burwood Hospital & School of Psychiatry, University of New South Wales, Burwood, Australia

[3]Royal Hospital for Women, Randwick, Australia

[*]all authors contributed equally

Corresponding Author:
Nicole Reilly, BPsyc, PhD
Centre for Health Service Development
Australian Health Services Research Institute
University of Wollongong
Building 234 (iC Enterprise 1), Innovation Campus
Wollongong, 2522
Australia
Phone: 61 0242218045
Email: nreilly@uow.edu.au

Abstract

Background: *Mummatters* is a web-based health tool that allows women to self-assess the symptoms of depression and the presence of psychosocial risk factors throughout pregnancy and the postnatal period. It aims to increase women's awareness of their own symptoms or risk factors and their knowledge of the available support options, to encourage engagement with these support options (as appropriate), and to facilitate communication about emotional health issues between women and their health care providers.

Objective: The aim of this study is to report the uptake of *mummatters*; the sociodemographic and psychosocial risk profiles of a subsample of users; and the acceptability, credibility, perceived effect, and motivational appeal of the tool. The help-seeking behaviors of the subsample of users and barriers to help seeking were also examined.

Methods: *Mummatters* was launched in November 2016. Women who completed the *mummatters* baseline assessment were invited to complete a web-based follow-up survey 1 month later.

Results: A total of 2817 women downloaded and used *mummatters* between November 13, 2016, and May 22, 2018, and 140 women participated in the follow-up study. Approximately half of these women (51%; 72/140) were *Whooley positive* (possible depression), and 43% (60/140) had an elevated psychosocial risk score on the Antenatal Risk Questionnaire. *Mummatters* was rated favorably by pregnant and postnatal women in terms of its acceptability (94%-99%), credibility (93%-97%), appeal (78%-91%), and potential to affect a range of health behaviors specific to supporting emotional wellness during the perinatal period (78%-93%). *Whooley-positive* women were more likely to speak with their families than with a health care provider about their emotional health. Normalizing symptoms and stigma were key barriers to seeking help.

Conclusions: Although *mummatters* was rated positively by consumers, only 53% (19/36) to 61% (22/36) of women with possible depression reported speaking to their health care providers about their emotional health. There was a trend for more prominent barriers to seeking help among postnatal women than among pregnant women. Future studies that investigate whether social barriers to seeking help are greater once a woman has an infant are warranted. Such barriers potentially place these women at greater risk of remaining untreated, as the demands on them are greater.

(J Med Internet Res 2021;23(3):e18517) doi:10.2196/18517

KEYWORDS

pregnancy; postpartum; self-assessment; depression; risk

Introduction

Background

Overall, 1 in 7 women experience some form of mental health morbidity during pregnancy and the first postnatal year (the perinatal period) [1,2]. When left untreated, perinatal depression and anxiety may persist for years after birth and can affect not only the woman's capacity to parent but also the emotional well-being of the infant and other family members [3,4]. The importance of early detection and intervention in this susceptible population has been widely acknowledged [5-9].

The health service systems in place for routine maternity care in Australia have provided a unique opportunity to introduce perinatal mental health promotion, prevention, and early intervention programs. These programs and related clinical guidelines aim for the early identification of possible or probable illness, or risk of illness, and then to monitor or intervene as appropriate, with a view to improve maternal mental health outcomes [6,10]. However, there are disparities in access to these programs, with women who give birth in the private maternity sector, for example, being less likely to be assessed across various domains of psychosocial health during pregnancy or the postpartum period [11-13].

In response to this inequity of access, Bupa Australia collaborated with perinatal mental health and consumer teams to develop a web-based tool, *mummatters* [14], which allows women to self-assess for the symptoms of depression and the presence of psychosocial risk factors throughout the perinatal period. *mummatters* aims to increase women's awareness of their own symptoms or risk factors and their knowledge of the available support options, encourage engagement with these support options (as appropriate), and facilitate communication about emotional health issues between women and their health care providers.

Objective

The aim of this study was to report on the uptake of *mummatters*; the sociodemographic and psychosocial risk profiles of a subsample of users; and the acceptability, credibility, perceived effect, and motivational appeal of the tool to pregnant and postnatal women. The help-seeking behaviors of women who used *mummatters* and barriers to help seeking were also examined.

Methods

Mummatters Overview

Mummatters is a web-based tool that is available free of charge via the Bupa website [14]. Its web-based design allows access from a range of computing and mobile devices. Users can bookmark the website or save *mummatters* to the home page of their devices, where it appears as an icon. Internet access is required to use the features of the tool.

After downloading *mummatters*, women are invited to answer a small number of demographic questions (including current gestation or infant age and maternity care sector), followed by a baseline assessment comprising the *Whooley* questions [15,16] and the Antenatal Risk Questionnaire (ANRQ; and its postnatal equivalent) [17].

The *Whooley* depression case-finding questions are recommended for use in the perinatal period by the National Institute for Health and Care Excellence [8]. The 2-item questionnaire has been shown to have a high sensitivity (0.95; 95% CI 0.88-0.97) and modest specificity (0.65; 95% CI 0.56-0.74) [18]. The *Whooley* questions are as follows: (1) "During the past month, have you often been bothered by feeling down, depressed or hopeless?" and (2) "During the past month, have you often been bothered by having little interest or pleasure in doing things?" [15]. These depression-related questions are followed by a third question that is asked to women who responded "yes" to either of the 2 above-mentioned questions ("Is this something you feel you need or want help with?") [16]. Women were considered to be *Whooley positive* (possible depressive episode) if they answered "yes" to either/both questions 1 or 2 [8].

The ANRQ is a validated self-report measure that was developed by a panel of experts, based on evidence relating to salient risk factors associated with perinatal mental health disorders, particularly depression and anxiety, and on the face and construct validity of these factors. Its capacity to identify women at increased risk for these conditions has been demonstrated [17]. Although initially developed for the antenatal period, the ANRQ has been used during the postnatal period for research and clinical practice [19,20]. Australia's current clinical practice guidelines for mental health care in the perinatal period recommend the use of the ANRQ for the assessment of psychosocial risk [6].

A key feature of *mummatters* is its computer-based decision aid that combines responses to the Whooley questions and ANRQ to generate tailored follow-up messages and provide help-seeking information, as appropriate. For example, women who are *Whooley positive* automatically receive a message encouraging them to make an appointment to discuss their emotional health with a trusted health care professional. The tool also allows women to give permission for a letter addressed to their health care provider to be generated, which includes a summary of their results as well as full copies of their completed measures. Women are also given ready access to additional information and links to resources that aim to support them in actively looking after their emotional well-being. After the completion of the initial baseline assessment, *mummatters* sends monthly prompts for women (irrespective of their baseline scores) to complete follow-up assessments to monitor their emotional health and well-being. Women can also create an individualized wellness action plan and can opt to receive inspirational messages sent monthly via SMS or email.

Data Collection and Research Participants

There were 2 primary sources of data for this evaluation. The first was *mummatters* use data that women consented to being used for research purposes via a within-tool agreement. These use data included a unique identification number as well as deidentified demographic and clinical information provided during the initial pregnancy or postnatal assessment.

The second data source was a research-specific data set that required additional consent. All women who used *mummatters* at least once during pregnancy or postnatally and who indicated their willingness to be contacted about the study were emailed and invited to participate. To be eligible, women were also required to be currently living in Australia, have access to the internet, and be able to complete the measures in English. Eligible women who agreed to participate gave informed consent and completed the additional study measures via the web-based *Key Survey* (TM) platform. A reminder email was sent by the research team to women who did not complete the study measures within 1 week, with 2 further reminders sent at weekly intervals thereafter (up to a maximum of 3 reminders).

Participants completed questions relating to the acceptability, credibility, likeability, perceived effect, and motivational appeal of the tool. Participants were also asked about help-seeking behaviors in the previous month and barriers to help seeking. Where possible, these questions were replicated or modified from previous studies for methodological consistency [21,22]. These data were linked to the use data of the participants via their unique identification numbers.

Ethical Approval

The study was approved by the Human Research Ethics Committee (HREC) of St John of God Health Care (HREC reference number: 735).

Results

Sociodemographic and Psychosocial Profile of Mummatters Evaluation Participants

A total of 2817 women downloaded and completed the *mummatters* baseline measures between November 13, 2016, and May 22, 2018. Of these, 26.80% (755/2817) indicated their willingness to be contacted about the study and were emailed an information sheet, consent form, and link to the study measures approximately 4 weeks after indicating their expression of interest. Of these, 33.8% (255/755) women agreed to participate in the study; among these, 91 women dropped out of the survey immediately after indicating their consent (ie, before completing any of the research questions). Of the remaining 164 women, 140 had sufficient research data and were subsequently included in the analyses (ie, 5% (140/2817) of all *mummatters* users and 18.5% (140/755) of all women emailed about the research). The demographic profiles of the 140 women who participated are presented in Table 1.

There were no significant differences between pregnant or postnatal *mummatters* users who were and were not included in the study in terms of *Whooley-positive* status (antenatal: $\chi^2(2)=1.0$; $P=.60$ and postnatal: $\chi^2(2)=2.5$; $P=.29$), ANRQ total score (antenatal: $t_{1331}=-1.32$; $P=.19$ and postnatal: $t_{1369}=-0.05$; $P=.96$), private maternity sector (antenatal: $\chi^2(2)=1.1$; $P=.78$ and postnatal: $\chi^2(1)=0.7$; $P=.68$), gestation ($t_{1377}=0.34$; $P=.73$), or infant age ($t_{1433}=1.38$; $P=.17$) at baseline assessment.

The results of the *mummatters* baseline psychosocial assessment (Whooley questions and psychosocial risk questionnaire) for participants are summarized in Table 2. During pregnancy, 49% (n=36/73) of participants were *Whooley positive*; one or both of the Whooley questions were endorsed by 15% (11/73) and 34% (25/73) of participating women, respectively. In the postnatal period, 36 of 67 (54%) participants were *Whooley positive*, with 31% (21/67) of women endorsing one question and 22% (15/67) endorsing both questions. Across the perinatal period, women who endorsed both Whooley questions were significantly more likely to answer “yes” to the third Whooley *help* question than women who endorsed one question only (antenatal: $\chi^2(1)=5.7$; $P=.02$ and postnatal: $\chi^2(1)=5.2$; $P=.02$).

Table 1. Sociodemographic characteristics of mummatters evaluation participants.

Characteristic	Antenatal period (baseline; n=73)	Postnatal period (baseline; n=67)
Gestation or infant age at first *mummatters* baseline assessment (weeks)		
Mean (SD)	20.96 (11.19)	15.28 (24.11)
Range	4-40	1-178
Maternal age (years)		
Mean (SD)	32.97 (4.60)	32.70 (4.20)
Range	24-43	25-45
Partnered[a], n (%)	68 (96)	65 (99)
First child[b], n (%)	42 (62)	28 (44)
Australian born[b], n (%)	60 (88)	49 (78)
Maternity sector[c], n (%)		
Public	32 (44)	30 (46)
Private	37 (51)	36 (55)
Undecided	4 (6)	N/A[d]
Residential area[a], n (%)		
Metropolitan	51 (72)	43 (65)
Regional	14 (20)	20 (30)
Rural	6 (9)	3 (5)
Highest educational level[a], n (%)		
Highschool	6 (9)	4 (6)
TAFE[e] or diploma	20 (28)	15 (23)
Bachelor's degree	30 (42)	28 (42)
Postgraduate degree	15 (21)	19 (29)
Current employment status[a], n (%)		
On maternity leave	19 (27)	45 (68)
Full time	33 (47)	4 (6)
Part time	12 (17)	6 (9)
Unemployed	7 (10)	11 (17)
Income management[a,f], n (%)		
Difficult	17 (24)	26 (39)
Not difficult	54 (76)	40 (61)

[a]Excludes 2 antenatal women and 1 postnatal woman with missing data.

[b]Excludes 5 antenatal women and 4 postnatal women with missing data.

[c]Excludes 1 postnatal woman with missing data.

[d]N/A: not applicable.

[e]TAFE: Technical and Further Education.

[f]Responses to the question "How do you manage on the income you have available" were divided into difficult ("It is impossible," "It is difficult all of the time," and "It is difficult some of the time") or not difficult ("It is too bad" and "It is easy") responses.

Columns may not total 100% due to rounding.

Table 2. Psychosocial profile of mummatters evaluation participants.

Characteristic	Antenatal period (baseline; n=73)	Postnatal period (baseline; n=67)
Psychosocial risk questionnaire (ANRQ[a]) total score		
Mean (SD)	23.11(10.64)	23.09 (10.00)
Range	6-53	7-49
Whooley positive[b], n (%)	36 (49)	36 (54)
ANRQ score≥23, n (%)	30 (41)	30 (45)
ANRQ item, n (%)		
High trait anxiety (being a worrier)[c]	20 (27)	22 (33)
Perfectionistic traits[c]	19 (26)	14 (21)
Past sexual or physical abuse	17 (23)	9 (13)
Significant past mental health issues[d]	31 (22)	18 (27)
Major stressors in the last 12 months[e]	16 (22)	18 (27)
Emotional abuse in childhood	14 (19)	11 (16)
Growing up with emotionally unsupportive mother[c]	13 (18)	18 (27)
Lack of support with the baby[c]	7 (10)	12 (18)
Emotionally unsupportive partner[c] (or no partner)	6 (8)	4 (6)

[a]ANRQ: Antenatal Risk Questionnaire.

[b]Endorsed one or both of the two Whooley questions.

[c]Antenatal Risk Questionnaire; scaled items were dichotomized into low scoring (1-3) or high scoring (4 or more), consistent with the methodology used in previous research [23,24].

[d]Item considered endorsed if participants responded “yes” to depression or other past mental health problems for which professional help was sought or which significantly interfered with work and relationships (score of 4 or more).

[e]Item considered endorsed if participants responded “yes” to experiencing a major stressor in the previous 12 months which caused a significant degree of distress (score of 4 or more).

Overall, 33% (24/73) of women who completed the antenatal baseline assessment did not endorse any significant risk factors on the ANRQ, 26% (19/73) endorsed 1 significant risk factor, 15% (11/73) endorsed 2 significant risk factors, and 26% (19/73) had 3 or more significant risk factors. Postnatally, 19% (13/67), 24% (16/67), 31% (21/67), and 25% (17/67) of women endorsed none, 1, 2, and 3 or more risk factors on the ANRQ, respectively. A total of 41% (30/73) of antenatal participants and 45% (30/67) postnatal participants scored 23 or more, which is the cutoff score for the ANRQ. Women who score above the cutoff are considered to be experiencing a significant accumulation of risk factors that are associated with an increased risk of developing a clinical depression or anxiety disorder [17].

Participant Experience and Feedback

The mean time taken by women to complete the *mummatters* baseline questions was 4 minutes (antenatal women: mean 4.16 minutes, SD 11.27 minutes; postnatal women: mean 3.88 minutes, SD 8.48 minutes). Most pregnant and postpartum women rated *mummatters* favorably on a range of feedback parameters (Table 3). Of note, most pregnant and postnatal users regarded *mummatters* as acceptable (94.0%-98.6%), credible (93.2%-97.3%), appealing (78.1%-91.0%), and potentially helpful in affecting a range of health behaviors specific to supporting emotional wellness during the perinatal period (78.1%-92.5%).

Table 3. Participant agreement with feedback statements relating to the use of mummatters.

Feedback statement	Antenatal period (n=73), n (%)[a]	Postnatal period (n=67), n (%)[a]
Acceptability		
I felt comfortable completing questions about my emotional health and well-being using *mummatters*	72 (99)	63 (94)
Credibility		
The information I got from *mummatters* can be trusted	71 (97)	64 (96)
The information I got from *mummatters* was useful	68 (93)	64 (96)
Perceived effect		
The information in *mummatters* helped me better understand the importance of having good emotional health in the transition to motherhood	65 (89)	62 (93)
mummatters helped me learn about the symptoms of depression	57 (78)	61 (91)
mummatters helped me learn about some common risk factors for depression and anxiety during pregnancy and in the year after birth	63 (86)	61 (91)
mummatters will help me pay closer attention to my emotional health and well-being	58 (80)	54 (81)
mummatters would help me feel more comfortable in seeking support for emotional health issues during pregnancy and in the year after birth, if I needed it	62 (85)	55 (83)
mummatters would help reduce the stigma of seeking help for emotional health issues during pregnancy and in the year after birth, if I needed it	65 (89)	56 (84)
mummatters increased my awareness of additional resources for emotional well-being during pregnancy and in the year after birth	63 (86)	59 (88)
mummatters provides practical solutions to managing emotional health issues during pregnancy and in the year after birth	60 (82)	60 (90)
The report that I can download in *mummatters* would help me talk to my health care provider about my emotional well-being, if I needed to	63 (86)	58 (87)
The information provided in *mummatters* could help me manage my emotional well-being in the future	63 (86)	61 (91)
Motivational appeal		
I would use *mummatters* again	66 (90)	61 (91)
I would tell friends to use *mummatters*	57 (78)	57 (85)
Likeability		
It was easy to find the information I wanted in *mummatters*	66 (90)	59 (88)
The information I got from *mummatters* was relevant to me	66 (90)	62 (93)
Overall, the features of *mummatters* met my expectations	58 (80)	54 (81)

[a]Numbers and percentages indicate those who agreed or strongly agreed with each statement.

Help-Seeking Behaviors and Barriers to Help Seeking

The help-seeking behaviors of *Whooley-positive* women (ie, women endorsing one or both of the Whooley depression questions) in the month before completing the study survey as well as the barriers to help seeking are presented in Table 4. Overall, these women were the most likely to discuss their emotional health with their partners or family during both pregnancy (33/36, 92%) and the postnatal period (30/36, 83%). Other common sources of support were friends (56%-61%), health care providers (53%-61%), and books or print materials (56%-64%). Women were more likely to report using complementary therapies for their emotional health during pregnancy than after birth. Interestingly, only 23% (5/22) of *Whooley-positive* women who spoke with a health care provider during pregnancy took their *mummatters* report to the appointment with them, which decreased to 11% (2/19) in the postnatal period.

Table 4. Help-seeking behaviors and barriers to help seeking in the month before completing the study survey among antenatal and postnatal Whooley-positive women.

Help-seeking behaviors and barriers	Antenatal period (n=36), n (%)	Postnatal period (n=36), n (%)	*P* value[a]
Help-seeking behaviors in the month before completing the study survey			
Spoke to health care professional[b]	22 (61)	19 (53)	.47
Partner or family	33 (92)	30 (83)	.48
Friends	22 (61)	20 (56)	.54
Internet	13 (36)	15 (42)	.42
Books or print materials	20 (56)	23 (64)	.53
Lifestyle changes	9 (25)	15 (42)	.10
Complementary therapies (including supplements)	16 (44)	7 (19)	.04
Started or continued medication	7 (19)	3 (8)	.19
Phone helpline	0 (0)	3 (8)	.24
Day stay or residential parenting service	N/A[c]	3 (8)	N/A
Hospital emergency department or admission	0 (0)	2 (6)	.49
Barriers to help-seeking in the month before completing the study survey			
Did not think needed help	7 (19)	11 (30)	.31
Normalizing symptoms	11 (31)	17 (46)	.18
Not aware of services	6 (17)	6 (16)	.96
Would feel like a failure	6 (17)	12 (32)	.12
Fear of judgment	10 (28)	17 (46)	.11
Worried about side effects of treatment	6 (17)	4 (11)	.52
Could not afford it	6 (17)	3 (8)	.31
Could not arrange childcare or transport	0 (0)	4 (11)	.12

[a]Chi-square test was used when n is >5, and Fisher exact test was used when n is <5.

[b]Includes midwife, child health nurse, general practitioner, obstetrician, counselor, psychologist, and psychiatrist.

[c]N/A: not applicable.

The most common barriers to seeking additional help or support reported by *Whooley-positive* women were personal or social in nature. For example, 46% (17/36) of postnatal participants normalized their symptoms or feared that they would be negatively judged if they asked for help, and up to one-third of women reported that they would feel like a failure. The proportion of women reporting these effects of stigma was greater in the postnatal period than during pregnancy, although these differences were not statistically significant.

Discussion

Principal Findings

This study sought to report on the uptake of *mummatters* and to provide insights into the experience, psychosocial profile, and help-seeking behaviors of women who engaged with this free, web-based self-assessment tool. Approximately 3000 women downloaded *mummatters* in 18 months. The results demonstrated that the tool was positively appraised by both pregnant and postnatal users, with high levels of reported acceptability, credibility, likeability, perceived effect, and motivational appeal. Approximately half of the women who used *mummatters* had chosen private maternity care for their current pregnancy or birth (48%; 1358/2817), suggesting that the tool was reaching a population that was known to be less likely to be offered depression screening and psychosocial assessment as a routine component of their antenatal and postnatal care.

Existing Australian research has reported elevated ANRQ scores among 14%-32% of women [23-25] in the general perinatal population, and although there is no Australian comparison data available for the Whooley questions, previous community-based studies have reported *Whooley-positive* rates of between 10% and 51% [26,27]. In comparison, up to 45% of women in this study scored above the recommended clinical cutoff score on the ANRQ and up to 54% were *Whooley positive*, suggesting that women who are at greater risk of poorer emotional health or parenting outcomes or who are experiencing current symptoms of depression are using *mummatters* and finding it highly acceptable. In keeping with recent research, this pattern of results may also reflect that an anonymous web-based assessment makes it easier for women to give an honest account of how they are feeling [28,29]. Although depression screening and psychosocial assessment are largely acceptable to most

women and health providers [17,23,24,30], recent research has shown that women who are most likely to need mental health care during the perinatal period are also those least likely to be honest with their health care providers when responding to questions about their mental health [31].

It is well-established that fear and shame are significant factors in women's decisions to seek or accept help for mental health issues during the perinatal period [32]. Approximately half of the women normalized their symptoms and were particularly concerned with how others would judge them if they admitted that they were struggling emotionally with motherhood. The barriers presented by these stigmatizing beliefs remain despite general population community surveys showing a high rate of disagreement with negative stereotypes about depression and motherhood, including disagreeing with the view that women with postnatal depression are unable to be good mothers [33].

Up to 47% of *Whooley-positive* women in this study did not speak to a health care professional about how they were feeling in the month after completing the *mummatters* baseline assessment, and only a few (10.5%-22.7%) of those that did took the downloadable *mummatters* report to their appointment. Although this is in line with previous Australian research that shows that up to 50% of women do not seek help for emotional health issues during the perinatal period despite being identified as in need of additional support [34,35], there is a clear need to understand why the report was being underutilized and how this can be made more useful, particularly given that women had already taken a first step in seeking a way to evaluate their symptoms. Seeking women's consent to automatically send the reports from *mummatters* to a nominated health care provider is one possible response so that the onus to seek care does not solely lie on the women, whose symptoms may inherently make it difficult to seek support. However, this option assumes that women will have continuity of care, but this is not always possible in contexts where care is delivered across hospital maternity care systems and postnatal community–based primary care systems and between the public and private health care sectors. Gathering women's views on how this feature can be improved or made more acceptable to users was beyond the scope of this study but is critical to inform future updates of the tool. Consistent with other Australian research [22,35,36], this study showed that family and friends are key support options for many women during pregnancy and the postnatal period. This again highlights the critical importance of targeting partners, family, and social networks in community awareness campaigns and early intervention programs for perinatal mental health. However, such campaigns must be complemented by support and treatment approaches that are well resourced, available, and enhance timely access to appropriate follow-up care. Interestingly, both pregnant and postnatal women in this study were more likely to report seeking support or information about their emotional health from books or print materials than the internet. This was despite women already being engaged with *mummatters* as a web-based tool and despite the increasing availability of locally developed internet-based resources and interventions [37]. The feasibility of partnering with service providers to directly link women to evidence-based web-based and telehealth treatment programs in future iterations of *mummatters* is currently being explored.

Strengths and Limitations

This study has several limitations. The sample size was small; however, based on the limited comparison data available, participants were representative of all *mummatters* users. Although it was a self-selected sample, a greater than expected proportion of participants endorsed possible depression or a substantial number of psychosocial risk factors, predisposing them to developing a mental health episode. Thus, we were able to examine the concerns of more susceptible women more closely in terms of help-seeking behaviors and barriers.

Conclusions

Previous research has reported on the engagement and real-world clinical utility of web-based approaches to self-administered screening for mental health conditions [38-41]; however, such approaches have not been adequately evaluated in perinatal populations. This study provided insight into the profile, experience, and help-seeking behaviors of women who used *mummatters*, a freely available web-based tool, and our results will help inform the review and further development of the tool. Although *mummatters* was rated positively by consumers, only 53% (19/36) to 61% (22/36) of women with possible depression reported speaking to their health care provider about this. This was more notable (though not statistically significant) among postnatal women than among pregnant women, suggesting that the barriers to help seeking are greater once a woman has an infant. Such barriers potentially place these women at greater risk of remaining untreated, as the demands on them are greater. This warrants further investigation in future studies.

Although consumer-driven risk assessments and symptom checklists are becoming more readily available for perinatal women, the need to keep training health care providers to engage women and ask the right questions to start the conversation around emotional well-being remains imperative [42]. Multiple but complementary approaches may be necessary given the well-documented findings, supported by our study, that many women do not seek formal assistance from their health care providers even when they are encouraged to do so. Future research should also focus on whether engagement with self-assessment tools of this type, including frequency and duration of engagement, is associated with longer-term improvements in mental health and health-related quality of life.

Acknowledgments

The authors thank the women who generously gave their time to participate in this study. The authors gratefully acknowledge the infrastructure funding of St John of God Health Care and thank the clinical leadership, health and care solutions, and research

and partnerships divisions of Bupa Australia for their investment in *mummatters* and collaborative approach to research translation and knowledge exchange. NR also thanks Australian Rotary Health, the University of Newcastle (2018-2020), and the University of Wollongong (2020-2023) for postdoctoral fellowship funding support. The funding bodies had no involvement in the design of the study, analysis and interpretation of data, or writing of the manuscript.

Conflicts of Interest

None declared.

References

1. Gaynes BN, Gavin N, Meltzer-Brody S, Lohr KN, Swinson T, Gartlehner G, et al. Perinatal depression: prevalence, screening accuracy, and screening outcomes. Evid Rep Technol Assess (Summ) 2005 Feb(119):1-8. [Medline: 15760246]
2. Dennis C, Falah-Hassani K, Shiri R. Prevalence of antenatal and postnatal anxiety: systematic review and meta-analysis. Br J Psychiatry 2017 May;210(5):315-323. [doi: 10.1192/bjp.bp.116.187179] [Medline: 28302701]
3. Woolhouse H, Gartland D, Mensah F, Brown SJ. Maternal depression from early pregnancy to 4 years postpartum in a prospective pregnancy cohort study: implications for primary health care. BJOG 2015 Feb;122(3):312-321. [doi: 10.1111/1471-0528.12837] [Medline: 24844913]
4. Stein A, Pearson RM, Goodman SH, Rapa E, Rahman A, McCallum M, et al. Effects of perinatal mental disorders on the fetus and child. Lancet 2014 Nov 15;384(9956):1800-1819. [doi: 10.1016/S0140-6736(14)61277-0] [Medline: 25455250]
5. RANZCOG. Mental Health Care in the Perinatal Period (C-Obs 48). Melbourne: Royal Australian and New Zealand College of Obstetricians and Gynaecologists; 2018. URL: https://ranzcog.edu.au/statements-guidelines [accessed 2020-10-15]
6. Austin MP, Highet N. and the Expert Working Group. Mental Health Care in the Perinatal Period: Australian Clinical Practice Guideline. Melbourne: COPE: Centre of Perinatal Excellence; 2017. URL: https://cope.org.au/wp-content/uploads/2017/10/Final-COPE-Perinatal-Mental-Health-Guideline.pdf [accessed 2020-10-15]
7. DoH. Clinical Practice Guidelines: Pregnancy Care. Canberra: Department of Health, Australian Government; 2018. URL: https://www.health.gov.au/resources/pregnancy-care-guidelines [accessed 2020-10-15]
8. NICE. NICE Clinical Guideline 192. Antenatal and postnatal mental health: Clinical management and service guidance. London: National Institute for Health and Care Excellence; 2014. URL: https://www.nice.org.uk/guidance/cg192 [accessed 2020-10-15]
9. O'Connor E, Rossom RC, Henninger M, Groom HC, Burda BU. Primary Care Screening for and Treatment of Depression in Pregnant and Postpartum Women: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA 2016 Jan 26;315(4):388-406. [doi: 10.1001/jama.2015.18948] [Medline: 26813212]
10. DoHA. National Perinatal Depression Initiative 2008 - 2013. Canberra: Department of Health and Ageing, Australian Government URL: http://www.health.gov.au/internet/main/publishing.nsf/Content/mental-perinat [accessed 2020-10-15]
11. Reilly N, Harris S, Loxton D, Chojenta C, Forder P, Milgrom J, et al. Disparities in reported psychosocial assessment across public and private maternity settings: a national survey of women in Australia. BMC Public Health 2013 Jul 04;13:632 [FREE Full text] [doi: 10.1186/1471-2458-13-632] [Medline: 23826627]
12. Yelland J, Brown SJ. Asking women about mental health and social adversity in pregnancy: results of an Australian population-based survey. Birth 2014 Mar;41(1):79-87. [doi: 10.1111/birt.12083] [Medline: 24571172]
13. San Martin Porter MA, Betts K, Kisely S, Pecoraro G, Alati R. Screening for perinatal depression and predictors of underscreening: findings of the Born in Queensland study. Med J Aust 2019 Jan;210(1):32-37. [doi: 10.5694/mja2.12030] [Medline: 30636310]
14. Bupa Australia. mummatters. Melbourne: Bupa Australia Health Pty Ltd; 2016. URL: https://www.bupa.com.au/mummatters [accessed 2020-10-15]
15. Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression. Two questions are as good as many. J Gen Intern Med 1997 Jul;12(7):439-445 [FREE Full text] [Medline: 9229283]
16. Arroll B, Goodyear-Smith F, Kerse N, Fishman T, Gunn J. Effect of the addition of a "help" question to two screening questions on specificity for diagnosis of depression in general practice: diagnostic validity study. BMJ 2005 Oct 15;331(7521):884 [FREE Full text] [doi: 10.1136/bmj.38607.464537.7C] [Medline: 16166106]
17. Austin MP, Colton J, Priest S, Reilly N, Hadzi-Pavlovic D. The Antenatal Risk Questionnaire (ANRQ): acceptability and use for psychosocial risk assessment in the maternity setting. Women Birth 2013 Mar;26(1):17-25. [doi: 10.1016/j.wombi.2011.06.002] [Medline: 21764399]
18. Bosanquet K, Bailey D, Gilbody S, Harden M, Manea L, Nutbrown S, et al. Diagnostic accuracy of the Whooley questions for the identification of depression: a diagnostic meta-analysis. BMJ Open 2015 Dec 09;5(12):e008913 [FREE Full text] [doi: 10.1136/bmjopen-2015-008913] [Medline: 26656018]
19. Christl B, Reilly N, Smith M, Sims D, Chavasse F, Austin MP. The mental health of mothers of unsettled infants: is there value in routine psychosocial assessment in this context? Arch Womens Ment Health 2013 Oct;16(5):391-399. [doi: 10.1007/s00737-013-0360-0] [Medline: 23775393]

20. Christl B, Reilly N, Yin C, Austin MP. Clinical profile and outcomes of women admitted to a psychiatric mother-baby unit. Arch Womens Ment Health 2015 Dec;18(6):805-816. [doi: 10.1007/s00737-014-0492-x] [Medline: 25591925]
21. Begley C, Shegog R, Harding A, Goldsmith C, Hope O, Newmark M. Longitudinal feasibility of MINDSET: a clinic decision aid for epilepsy self-management. Epilepsy Behav 2015 Mar;44:143-150. [doi: 10.1016/j.yebeh.2014.12.031] [Medline: 25705825]
22. Reilly N, Harris S, Loxton D, Chojenta C, Forder P, Austin MP. The impact of routine assessment of past or current mental health on help-seeking in the perinatal period. Women Birth 2014 Dec;27(4):e20-e27. [doi: 10.1016/j.wombi.2014.07.003] [Medline: 25154772]
23. Reilly N, Yin C, Monterosso L, Bradshaw S, Neale K, Harrison B, et al. Identifying psychosocial risk among mothers in an Australian private maternity setting: A pilot study. Aust N Z J Obstet Gynaecol 2015 Oct;55(5):453-458. [doi: 10.1111/ajo.12370] [Medline: 26222838]
24. Kalra H, Reilly N, Austin MP. An evaluation of routine antenatal depression screening and psychosocial assessment in a regional private maternity setting in Australia. Aust N Z J Obstet Gynaecol 2018 Dec;58(6):629-635. [doi: 10.1111/ajo.12777] [Medline: 29380345]
25. Priest SR, Austin MP, Barnett BB, Buist A. A psychosocial risk assessment model (PRAM) for use with pregnant and postpartum women in primary care settings. Arch Womens Ment Health 2008 Dec;11(5-6):307-317. [doi: 10.1007/s00737-008-0028-3] [Medline: 18726142]
26. Howard LM, Ryan EG, Trevillion K, Anderson F, Bick D, Bye A, et al. Accuracy of the Whooley questions and the Edinburgh Postnatal Depression Scale in identifying depression and other mental disorders in early pregnancy. Br J Psychiatry 2018 Jan;212(1):50-56 [FREE Full text] [doi: 10.1192/bjp.2017.9] [Medline: 29433610]
27. van Heyningen T, Honikman S, Tomlinson M, Field S, Myer L. Comparison of mental health screening tools for detecting antenatal depression and anxiety disorders in South African women. PLoS One 2018;13(4):e0193697 [FREE Full text] [doi: 10.1371/journal.pone.0193697] [Medline: 29668725]
28. Drake E, Howard E, Kinsey E. Online screening and referral for postpartum depression: an exploratory study. Community Ment Health J 2014 Apr;50(3):305-311 [FREE Full text] [doi: 10.1007/s10597-012-9573-3] [Medline: 23283485]
29. Kingston DE, Biringer A, Toosi A, Heaman MI, Lasiuk GC, McDonald SW, et al. Disclosure during prenatal mental health screening. J Affect Disord 2015 Nov 1;186:90-94. [doi: 10.1016/j.jad.2015.07.003] [Medline: 26233318]
30. Buist A, O'Mahen H, Rooney R. Acceptability, attitudes and overcoming stigma. In: Gemmill A, Milgrom J, editors. Identifying Perinatal Depression and Anxiety: Evidence-Based Practice in Screening, Psychosocial Assessment and Management. Chichester, West Sussex: John Wiley & Sons, Ltd; 2015:51-62.
31. Forder PM, Rich J, Harris S, Chojenta C, Reilly N, Austin MP, et al. Honesty and comfort levels in mothers when screened for perinatal depression and anxiety. Women Birth 2020 Mar;33(2):e142-e150. [doi: 10.1016/j.wombi.2019.04.001] [Medline: 31133524]
32. Dennis C, Chung-Lee L. Postpartum depression help-seeking barriers and maternal treatment preferences: a qualitative systematic review. Birth 2006 Dec;33(4):323-331. [doi: 10.1111/j.1523-536X.2006.00130.x] [Medline: 17150072]
33. Smith T, Gemmill AW, Milgrom J. Perinatal anxiety and depression: Awareness and attitudes in Australia. Int J Soc Psychiatry 2019 Aug;65(5):378-387. [doi: 10.1177/0020764019852656] [Medline: 31148499]
34. Woolhouse H, Brown S, Krastev A, Perlen S, Gunn J. Seeking help for anxiety and depression after childbirth: results of the Maternal Health Study. Arch Womens Ment Health 2009 Apr;12(2):75-83. [doi: 10.1007/s00737-009-0049-6] [Medline: 19214705]
35. Reay R, Matthey S, Ellwood D, Scott M. Long-term outcomes of participants in a perinatal depression early detection program. J Affect Disord 2011 Mar;129(1-3):94-103. [doi: 10.1016/j.jad.2010.07.035] [Medline: 20800898]
36. Rowe HJ, Holton S, Fisher J. Postpartum emotional support: a qualitative study of women's and men's anticipated needs and preferred sources. Aust J Prim Health 2013;19(1):46-52. [doi: 10.1071/PY11117] [Medline: 22951012]
37. PDeC. MumSpace: Support for the emotional health of new mums, at every step.: Perinatal Depression e-Consortium (PDeC); 2018. URL: https://www.mumspace.com.au/ [accessed 2020-10-15]
38. Parker G, Blanch B, Fletcher K, Greenfield L. Evaluating the utility of a bipolar self-test screening measure. Australas Psychiatry 2011 Aug;19(4):317-320. [doi: 10.3109/10398562.2011.578642] [Medline: 21879866]
39. Parker G, Fletcher K. The 'real world' utility of a web-based bipolar disorder screening measure: a replication study. J Affect Disord 2013 Sep 05;150(2):276-283. [doi: 10.1016/j.jad.2013.04.006] [Medline: 23701751]
40. Parker G, Fletcher K, Blanch B, Greenfield L. Take-up and profile of individuals accessing a web-based bipolar self-test screening measure. J Affect Disord 2012 Apr;138(1-2):117-122. [doi: 10.1016/j.jad.2011.12.013] [Medline: 22284017]
41. Parker G, Fletcher K, Blanch B, Greenfield L. The 'real world' utility of a web-based bipolar disorder screening measure. Acta Psychiatr Scand 2013 May;127(5):373-380. [doi: 10.1111/acps.12017] [Medline: 23039131]
42. Kingston D, Biringer A, Veldhuyzen van Zanten S, Giallo R, McDonald S, MacQueen G, et al. Pregnant Women's Perceptions of the Risks and Benefits of Disclosure During Web-Based Mental Health E-Screening Versus Paper-Based Screening: Randomized Controlled Trial. JMIR Ment Health 2017 Oct 20;4(4):e42 [FREE Full text] [doi: 10.2196/mental.6888] [Medline: 29054833]

Abbreviations

ANRQ: Antenatal Risk Questionnaire
HREC: Human Research Ethics Committee

Edited by R Kukafka; submitted 02.03.20; peer-reviewed by N Hussain-Shamsy, X He, F Dark; comments to author 29.06.20; revised version received 20.08.20; accepted 21.12.20; published 26.03.21.

Please cite as:
Reilly N, Austin MP
Attitudes and Engagement of Pregnant and Postnatal Women With a Web-Based Emotional Health Tool (Mummatters): Cross-sectional Study
J Med Internet Res 2021;23(3):e18517
URL: https://www.jmir.org/2021/3/e18517
doi:10.2196/18517
PMID:33769302

Original Paper

Blended Acceptance and Commitment Therapy Versus Face-to-face Cognitive Behavioral Therapy for Older Adults With Anxiety Symptoms in Primary Care: Pragmatic Single-blind Cluster Randomized Trial

Maartje Witlox[1]; Nadia Garnefski[1]; Vivian Kraaij[1]; Margot W M de Waal[2]; Filip Smit[3,4,5]; Ernst Bohlmeijer[6]; Philip Spinhoven[1]

[1]Faculty of Social and Behavioural Sciences, Section of Clinical Psychology, Institute of Psychology, Leiden University, Leiden, Netherlands
[2]Department Public Health & Primary Care, Leiden University Medical Center, Leiden, Netherlands
[3]Department of Mental Health & Prevention, Netherlands Institute of Mental Health and Addiction, Utrecht, Netherlands
[4]Department of Clinical, Neuro and Developmental Psychology, Vrije Universiteit, Amsterdam, Netherlands
[5]Department of Epidemiology & Biostatistics, Amsterdam University Medical Centers, VU University Medical Center, Amsterdam, Netherlands
[6]Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands

Corresponding Author:
Maartje Witlox
Faculty of Social and Behavioural Sciences
Section of Clinical Psychology, Institute of Psychology
Leiden University
Wassenaarseweg 52
Leiden, 2333 AK
Netherlands
Phone: 31 0611858085
Email: m.witlox@fsw.leidenuniv.nl

Abstract

Background: Anxiety symptoms in older adults are prevalent and disabling but often go untreated. Most trials on psychological interventions for anxiety in later life have examined the effectiveness of face-to-face cognitive behavioral therapy (CBT). To bridge the current treatment gap, other treatment approaches and delivery formats should also be evaluated.

Objective: This study is the first to examine the effectiveness of a brief blended acceptance and commitment therapy (ACT) intervention for older adults with anxiety symptoms, compared with a face-to-face CBT intervention.

Methods: Adults aged between 55-75 years (n=314) with mild to moderately severe anxiety symptoms were recruited from general practices and cluster randomized to either blended ACT or face-to-face CBT. Assessments were performed at baseline (T0), posttreatment (T1), and at 6- and 12-month follow-ups (T2 and T3, respectively). The primary outcome was anxiety symptom severity (Generalized Anxiety Disorder-7). Secondary outcomes were positive mental health, depression symptom severity, functional impairment, presence of Diagnostic and Statistical Manual of Mental Disorders V anxiety disorders, and treatment satisfaction.

Results: Conditions did not differ significantly regarding changes in anxiety symptom severity during the study period (T0-T1: B=.18, *P*=.73; T1-T2: B=−.63, *P*=.26; T1-T3: B=−.33, *P*=.59). Large reductions in anxiety symptom severity (Cohen *d*≥0.96) were found in both conditions post treatment, and these were maintained at the 12-month follow-up. The rates of clinically significant changes in anxiety symptoms were also not different for the blended ACT group and CBT group (χ^2_1=0.2, *P*=.68). Regarding secondary outcomes, long-term effects on positive mental health were significantly stronger in the blended ACT group (B=.27, *P*=.03, Cohen *d*=0.29), and treatment satisfaction was significantly higher for blended ACT than CBT (B=3.19, *P*<.001, Cohen *d*=0.78). No other differences between the conditions were observed in the secondary outcomes.

Conclusions: The results show that blended ACT is a valuable treatment alternative to CBT for anxiety in later life.

Trial Registration: Netherlands Trial Register TRIAL NL6131 (NTR6270); https://www.trialregister.nl/trial/6131

(J Med Internet Res 2021;23(3):e24366) doi:10.2196/24366

KEYWORDS

acceptance and commitment therapy; anxiety; older adults; internet interventions; cognitive behavioral therapy

Introduction

Background

Anxiety is among the most common mental health problems in older adults, with prevalence estimates for anxiety disorders ranging up to 15% [1-3]. When also considering the presence of anxiety symptoms that do not meet the diagnostic criteria for a disorder (so-called subclinical or subthreshold anxiety), estimates range between 15% and 52% [1,3]. Both anxiety disorders and subclinical anxiety in older adults are associated with limited physical and social activities, impairments in self-care, decreased well-being, comorbid depressive symptomatology, somatic problems, and increased use of benzodiazepines [4-6]. Despite the repeatedly demonstrated negative impact of anxiety in later life, only a small proportion of anxious older adults receive adequate psychological help [7-9]. This treatment gap is worrying as untreated anxiety symptoms in older adults tend to be chronic and aggravated over time [10].

The current scientific literature on psychological interventions for anxiety in later life is limited with regard to both well-evaluated treatment approaches and the precise types of anxiety they target. The large majority of trials in anxious older adults have investigated face-to-face cognitive behavioral therapy (CBT) for Generalized Anxiety Disorder (GAD). In the most recent meta-analysis on CBT for anxiety disorders in older adults that concluded CBT to be an effective treatment, 7 of the 12 included studies focused on GAD [11]. In recent years, researchers' focus has shifted a little to web-based and blended CBT interventions as treatment for anxiety in later life. To date, studies in older adults with heterogeneous anxiety symptomatology have found web-based CBT modules combined with guidance from a clinician to be effective in reducing symptom severity [12-15]. These results are promising, as scalable (partly) web-based interventions might be invaluable in bridging the current treatment gap in a cost-effective way.

As CBT is the only treatment that has been systematically studied and most studies thus far confirm its effectiveness, many clinical guidelines refer to it as the preferred treatment for older adults with anxiety [16-18]. However, to move the field forward and improve treatment of anxiety in later life, alternative treatment options should also be evaluated, because in most studies with active control conditions, effect sizes favoring CBT were small [11], and some evidence suggests that older adults benefit less from CBT for anxiety than younger adults [11,19,20]. It has been hypothesized that the cognitive aspects of challenging negative thoughts could be especially problematic for older adults [20]. Unfortunately, no high-quality studies on other treatment approaches have yet been published.

Acceptance and commitment therapy (ACT), a promising alternative to CBT, has been found to be effective in reducing anxiety symptoms in general adult samples, both in face-to-face and (partly) web-based formats [21,22]. Contrary to CBT, which focuses on re-evaluating cognitions and changing safety behavior and avoidance to achieve decreased levels of anxiety, ACT promotes acceptance-based emotion regulation and valued engagement in life [23]. ACT ultimately aims to increase psychological flexibility: the ability to fully and openly experience the present moment, including the negative aspects, and to behave in accordance with personal values [23]. It has been recognized as a treatment that explicitly aligns with the understanding of mental health as not only the absence of disease and illness but also the presence of the so-called positive mental health [24-26].

ACT might be especially suitable for older adults because its focus on stimulating acceptance and value-based action is consistent with age-related changes in emotion regulation and behavior. Reorientation on personal values and associated behavior change [27,28], present moment awareness, and willingness to experience and accept negative emotions have all been found to increase with age [29,30]. As some studies suggest that treatment is more effective when it draws upon a patient's strengths rather than remediating their shortcomings [31,32], ACT holds promise as a particularly suitable treatment approach for older adults. Another argument for ACT as a treatment option for anxiety in later life is its transdiagnostic focus on increasing psychological flexibility. Low levels of psychological flexibility have been related to both anxiety and depression symptoms [33], which often co-occur in older adults. Although ACT seems to be a promising treatment option for older adults with anxiety, so far only one pilot study that examined face-to-face ACT for late-life GAD has been published. None of the participants dropped out and worry and depression scores improved [28], leading the authors to conclude that ACT warrants a large-scale evaluation in anxious older adults.

Objectives

This trial aims to advance evidence-based treatment of anxiety in later life by evaluating the short- and long-term effectiveness of an ACT intervention in a large sample of older adults with anxiety symptoms. Specifically, we will evaluate a blended ACT intervention, because scalable internet-based interventions could be crucial in bridging the treatment gap in anxious older adults and should therefore be thoroughly evaluated. Furthermore, the low-threshold nature and easy accessibility of internet-based interventions might be especially appealing to older adults, who are known to experience barriers in seeking and receiving regular psychological treatment [9]. The blended ACT intervention will be compared with a face-to-face CBT intervention, which can be considered treatment as usual in the study setting [34-36]. As the ACT approach aligns with age-related changes in emotion regulation and behavior, we expect the ACT intervention to be more effective than CBT. In addition to the effect on the primary outcome anxiety symptom severity, the effects of interventions on positive mental health,

depressive symptoms, functional impairment, presence of anxiety disorders, and treatment satisfaction will be evaluated. As this study is the first large-scale trial into an ACT intervention for anxiety in later life, the results will offer valuable new insights into how the large and currently underserved group of older adults with anxiety symptoms can be treated.

Methods

Design

The study was registered in the Netherlands Trial Register (NL6131; NTR6270) and approved by the Medical Ethics Committee of Leiden University Medical Center (P16.248). A detailed description of the study protocol has been published [37].

The study was designed as a pragmatic, single-blind cluster, randomized controlled trial with measurements at baseline (T0) and follow-ups at 3, 6, and 12 months (T1, T2, T3, respectively) postbaseline. Randomization took place at the level of mental health counselors working in general practices, creating clusters of participants who received treatment from the same counselor. Power analysis showed that to detect a between-group difference on the Generalized Anxiety Disorder-7 (GAD-7) at posttreatment with a medium effect size (Cohen d=0.45), a 2-tailed α of .05, and a power of 0.80 [38], posttreatment data of 180 participants were required. Anticipating a dropout rate of 25%, 240 participants (36 counselors) were included at baseline.

The block-randomization table (blocks of 4) was created by an independent researcher using the *R* software [38] and was concealed from the main researcher. If 4 mental health counselors had registered for participation, the main researcher received their allocation from an independent researcher. After a mental health counselor was informed about their randomization status and had received training in the treatment they were allocated to provide to study participants, recruitment of participants from the general practice that employed the counselor started.

Research assistants (Master's students or graduates in clinical psychology) who conducted telephonic diagnostic interviews as part of the assessments were blinded to the participants' treatment assignments. The main researcher, mental health counselors, and participants were not blinded to treatment allocation. Study participants were not informed whether the intervention they received was the experimental or the active control condition. To prevent selection bias, potential participants were not informed about the randomization status of the mental health counselor in their general practice (ie, the intervention they would receive if they participated in the study) until they had given their informed consent and completed the baseline assessment.

Study Setting: General Practices

The treatment was provided by mental health counselors working in general practices in the Netherlands. Since 2008, general practices in the Netherlands have employed mental health counselors in response to the increasing demand for psychological treatment and the limited capacity of mental health care institutions [34]. The counselors offer brief psychological interventions to patients with mild to moderately severe symptomatology in the easily accessible environment of general practices.

General practices were recruited by sending information and invitation letters to practices in the networks of Leiden University and Leiden University Medical Centre. Furthermore, study information was distributed through messages in relevant newsletters and online forums. When a general practice agreed upon study participation, employees of the practice were asked to distribute the information among their professional networks. A total of 38 general practices were recruited. These practices were located in villages (n=10), towns (n=11), and cities (n=17) throughout the Netherlands, in 9 out of 12 provinces. The practices employed a total of 40 mental health counselors, who were randomized to provide study participants with either blended ACT (n=20) or face-to-face CBT (n=20). In total, 36 practices employed one mental health counselor and 2 practices employed 2 counselors each. Regarding the counselors' educational background, most were psychologists (n=13), social psychiatric nurses (n=14), or social workers (n=5). Two counselors were trained as system therapists, and the other 6 had different educational backgrounds. The number of years of experience in providing individual psychological treatment ranged from 3 to 42, with a median of 16 years.

Participants

Individuals aged between 55-75 years with mild to moderately severe anxiety symptoms (GAD-7 between 5 and 15 [39]) were eligible for participation. Mastery of the Dutch language, internet access, and motivation to spend 2.5 h per week on the intervention were also required. Exclusion criteria were severe cognitive impairment or unstable severe medical conditions (according to the medical record at the general practice); very mild or severe anxiety symptoms (GAD-7 score <5 or >15 [39]); severe depressive symptomatology (Patient Health Questionnaire-9 [PHQ-9] score≥20 [40]), psychological or psychopharmacological treatment within the last 3 months, with the exception of stable benzodiazepine or selective serotonin reuptake inhibitor use; severe functional impairment (score≥8 on 2 or 3 Sheehan Disability Scale (SDS) domains [41]), high suicide risk (Mini-International Neuropsychiatric Interview Plus [MINI-Plus]) [42]; substance use disorder (MINI-Plus); lifetime diagnosis of bipolar disorder or schizophrenia (medical record or MINI-Plus).

Procedure

Patients (aged between 55 and 75 years) from participating general practices were sent a letter containing information about anxiety symptoms, the aim and design of the study, and an invitation to participate. A data manager from the Leiden University Medical Center assisted general practitioners (GPs) in preparing and sending the letters in accordance with Dutch privacy legislation. Patients whose medical records mentioned a lifetime diagnosis of bipolar disorder or schizophrenia, severe unstable medical conditions, or severe cognitive impairment did not receive an invitation letter. GPs could also exclude patients from the mailing list for other reasons (eg, social

circumstances or language barriers) and had to give written approval of the final mailing list.

The information or invitation letters refer people to the study website for detailed information about the trial and to register for participation. After registration, they were screened using web-based questionnaires (assessing anxiety severity [GAD-7], depression severity [PHQ-9], mastery of Dutch, and motivation for treatment) and by a telephone interview (assessing medication use, functional impairment [SDS], and presence of psychiatric disorders [MINI-Plus]). If excluded for the presence of severe symptomatology, people were referred to their GP to discuss other treatment options. Web-based informed consent was obtained from all eligible participants before they completed the web-based baseline questionnaire. After this, the main researcher informed the included participants about the intervention they would receive and updated the general practice about the inclusion.

Participants completed 4 assessments (T0, T1, T2, and T3). Assessments mainly consisted of web-based self-report questionnaires. Assessments at T0, T1, and T3 were complemented by telephone interviews conducted by trained research assistants.

Treatments

Blended ACT

Participants in the blended ACT condition were given access to the web-based ACT-module *Living to the Full* and attended 4 face-to-face sessions with their mental health counselor at the general practice. The *Living to the Full* module consisted of 9 lessons to be completed in 9 to 12 weeks. This module (an adaptation of the similarly titled self-help book [43,44]) was proven effective in reducing distress and depression in earlier studies [45,46]. The web-based module could be accessed using computers and mobile devices. To complete the lessons in time, the participants were required to spend 15 minutes to 30 minutes on the module each day. The module consisted of 3 phases, each comprising 3 lessons. In the first phase, participants explored the negative consequences of their attempts to control or reduce their unwanted feelings or thoughts and were introduced to the idea of shifting their attitude toward their internal experiences from controlling to accepting. The next 3 lessons provided them with tools to be more accepting of their (unwanted) internal experiences: exercises focused on noticing thoughts and feelings without judgment and conceptualizing the self as the consciousness that notices internal experiences, instead of the content of these experiences. The last phase of the module focused on identifying core values and taking the first step toward living in accordance with these.

The authors of *Living to the Full* developed a treatment protocol for the 4 face-to-face sessions with the mental health counselor at the general practice. In the first session, the participants' complaints were inventoried and a web-based program was introduced. After this session, the participants were emailed their log-in credentials and could access the web-based module. The subsequent 3 lessons each connected to 1 of the 3 phases in the module and served to repeat key exercises, increase motivation, evaluate progress, and discuss potential problems. Mental health counselors could monitor the progress of their clients in the web-based module: they could see their answers to the exercises and the amount of time they spent on the module but could not provide web-based feedback.

Treatment-As-Usual: Face-to-face CBT

Participants in the treatment-as-usual group received a protocolized CBT intervention, consisting of 4 face-to-face sessions over a period of 9 to 12 weeks. In addition, participants were given homework exercises that required 15 to 30 min per day (ie, a similar time investment as the blended ACT intervention). The treatment protocol was developed by NG, MW, VK, and PS. It consisted of a manual with 12 different worksheets containing psychoeducation and CBT exercises. The main worksheets focused on thinking errors and avoidance behaviors. Other worksheets addressed specific forms of anxiety (eg, worrying, panic, social anxiety) or common consequences of anxiety (eg, sleep disturbances, muscle tension). On the basis of the intake and goal formulation during the first session, counselors and participants agreed upon which worksheets to use. In Session 2 and 3, the mental health counselor and participant discussed and repeated homework exercises, evaluated progress, and discussed potential problems, and the counselor aimed to increase the participants' motivation to continue with the intervention. The last session was dedicated to formulating a relapse prevention plan.

Mental health counselors received a 6-h long in-person training on working with the treatment protocol for their allocated treatment.

Measures

Table 1 presents an overview of the instruments used per measurement moment. Anxiety symptom severity was assessed using the GAD-7 (total scores 0-21), with higher scores indicating higher symptom severity [39]. Positive mental health was measured using the Mental Health Continuum-Short Form (MHC-SF; total scores: (range 0-5) were obtained by averaging the sum scores of the 14 6-point items, with higher scores indicating higher levels of positive mental health [47]). Depressive symptoms were assessed using the PHQ-9 (total score 0-27; higher scores reflect higher symptom severity [40]). The SDS [41] assessed functional impairment in the domains of work, social life, and family life (scores in each domain range 0-10, higher scores reflecting more impairment). The presence of current GAD, panic disorder, agoraphobia, specific phobia, social phobia, obsessive-compulsive disorder, posttraumatic stress disorder, and illness anxiety disorder according to DSM-V criteria was assessed using the MINI-Plus [42]. Treatment satisfaction was assessed using the Client Satisfaction Questionnaire-8 (total scores 0-32; higher scores indicate higher satisfaction [48]). To assess treatment integrity, mental health counselors, after every session, indicated how closely they had followed the treatment protocol on a checklist with all the elements the protocol prescribed for the sessions. Secondary outcomes not reported in this article were mindfulness, experiential avoidance, cognitive emotion regulation, medical costs, and quality of life. These outcomes will be used in subsequent studies on moderator, mediator, and cost-effectiveness analyses.

Table 1. Instruments per measurement moment.

Instrument	Screening	T0[a]	T1[b] (3 month)	T2 (6 month)	T3 (12 month)
Anxiety symptom severity (GAD-7[c])	✓[d]	✓	✓	✓	✓
Positive mental health (MHC-SF[e])	—[f]	✓	✓	✓	✓
Depression symptom severity (PHQ-9[g])	✓	✓	✓	✓	✓
Presence of psychiatric disorders[h] (MINI-Plus[i])	✓	✓	✓	—	✓
Functional impairment[h] (SDS[j])	✓	✓	✓	—	✓
Treatment satisfaction (CSQ-8[k])	—	—	✓	—	—

[a]T0: baseline.

[b]T1: posttreatment.

[c]GAD-7: Generalized Anxiety Disorder-7.

[d]Indicates that the instrument was used at the specified measurement moment.

[e]MHC-SF: Mental Health Continuum-Short Form.

[f]Indicates that the instrument was not used at the specified measurement moment.

[g]PHQ-9: Patient Health Questionnaire-9.

[h]Assessed during telephone interviews by trained research assistants. Scores on these measures obtained during screening are analyzed as part of baseline.

[i]MINI-Plus: Mini-International Neuropsychiatric Interview-Plus.

[j]SDS: Sheehan Disability Scale.

[k]CSQ-8: Client Satisfaction Questionnaire-8.

Statistical Analyses

Statistical analyses were performed using the *R* software [38]. The differences between conditions over time on continuous outcomes were examined using linear mixed models. The time variable was recoded into 3 contrasts: T0-T1 (baseline to posttreatment), T1-T2 (posttreatment to 6-month follow-up), and T1-T3 (posttreatment to 12-month follow-up). Functional impairment was not assessed at T2; therefore, these analyses included 2 contrasts (T0-T1 and T1-T3). The condition variable was effect-coded (CBT=−0.5, ACT=0.5) to ensure that the coefficients for the time variables reflected true main effects. Time, condition, and their interaction were included as fixed effects. Random intercepts were included at the participant level and mental health counselor level. Random slopes for time were included for mental health counselors but not for participants, as this would result in more parameters than observations. Treatment satisfaction was only assessed at T1, so this model included no time effects and only a random intercept at the counselor level. For this model, the condition was dummy coded (CBT=0, ACT=1).

Mixed effects logistic regression was used to examine if proportions of participants that changed from *anxiety disorder* to *no anxiety disorder*—and vice versa—differed between groups. A total of 4 separate models were created to examine the differences between the conditions at T1 and T3 for participants without an anxiety disorder. All mixed models were fitted to the data with the maximum likelihood. This method does not replace or impute missing values but uses all observed data to estimate the value of a population parameter by determining the value that maximizes the likelihood function [49].

Cohen *d* was used as the effect size for continuous outcomes and was calculated using mixed model estimated means and observed SD [50]. Cohen *d* values were interpreted as very small (<0.20), small (0.20-0.50), medium (0.50-0.80), or large (>0.80) [51]. Odds ratios were used as effect sizes for between-group differences on the binary outcome and were classified as small (1.49-3.45), medium (3.45-9), and large (>9) [52].

For participants with a GAD-7 posttreatment score, a reliable change index (RCI) was calculated by dividing the difference between baseline and posttreatment scores by the standard error of difference (SED) [52]. The test-retest reliability of the GAD-7 (0.83) was used to calculate the SED [39]. RCI values lower than −1.96 indicate reliable symptom improvement, and values over 1.96 denote deterioration [53]. Recovery was operationalized as a posttreatment score below the cut-off for moderately severe anxiety symptoms (GAD-7<10 [40]) for participants who scored above this cut-off at baseline. Participants with both reliable improvement and recovery met the criteria for clinically significant changes [53]. The proportions of participants with reliable improvement, deterioration, and clinically significant change in both groups were compared using the χ^2 test.

In addition to intention-to-treat (ITT) analyses, per-protocol (PP) analyses were also conducted. For both groups, PP treatment was defined as attending 3 or 4 (75% or more of the allocated treatment) of the face-to-face sessions.

Results

Participants

Figure 1 presents the flowchart of the participants. From November 2017 to March 2019, 35,820 invitation letters were sent. A total of 683 people were screened, of whom 314 were included: 150 in the blended ACT group and 164 in the CBT group. Table 2 shows the demographic and clinical characteristics of the participants. A total of 13 participants in the ACT group and 17 in the CBT group did not start the treatment, as they did not show up for the first appointment and later indicated that they wanted to stop their participation or were not reachable by phone and email to discuss further participation. At T1, 70.7% (222/314) of the participants completed the web-based questionnaire (ACT 101/150, 67.3%, CBT 121/164, 73.8%); at T2, 63.7% (200/314; ACT 88/150, 58.6%, CBT 112/164, 68.3%), and at T3, 56.7% (178/314; ACT 82/150, 55%, CBT 96/164, 59%). Telephone interviews at T1 and T3 were completed by 66% (208/314; ACT 92/150, 61.3%, CBT 115/164, 70.1%) and 44.6% (140/314; ACT 69/150, 46.0%; CBT 71/164, 43.3%), respectively.

Figure 1. Flowchart of study participants. ACT: acceptance and commitment therapy; CBT: cognitive behavioral therapy.

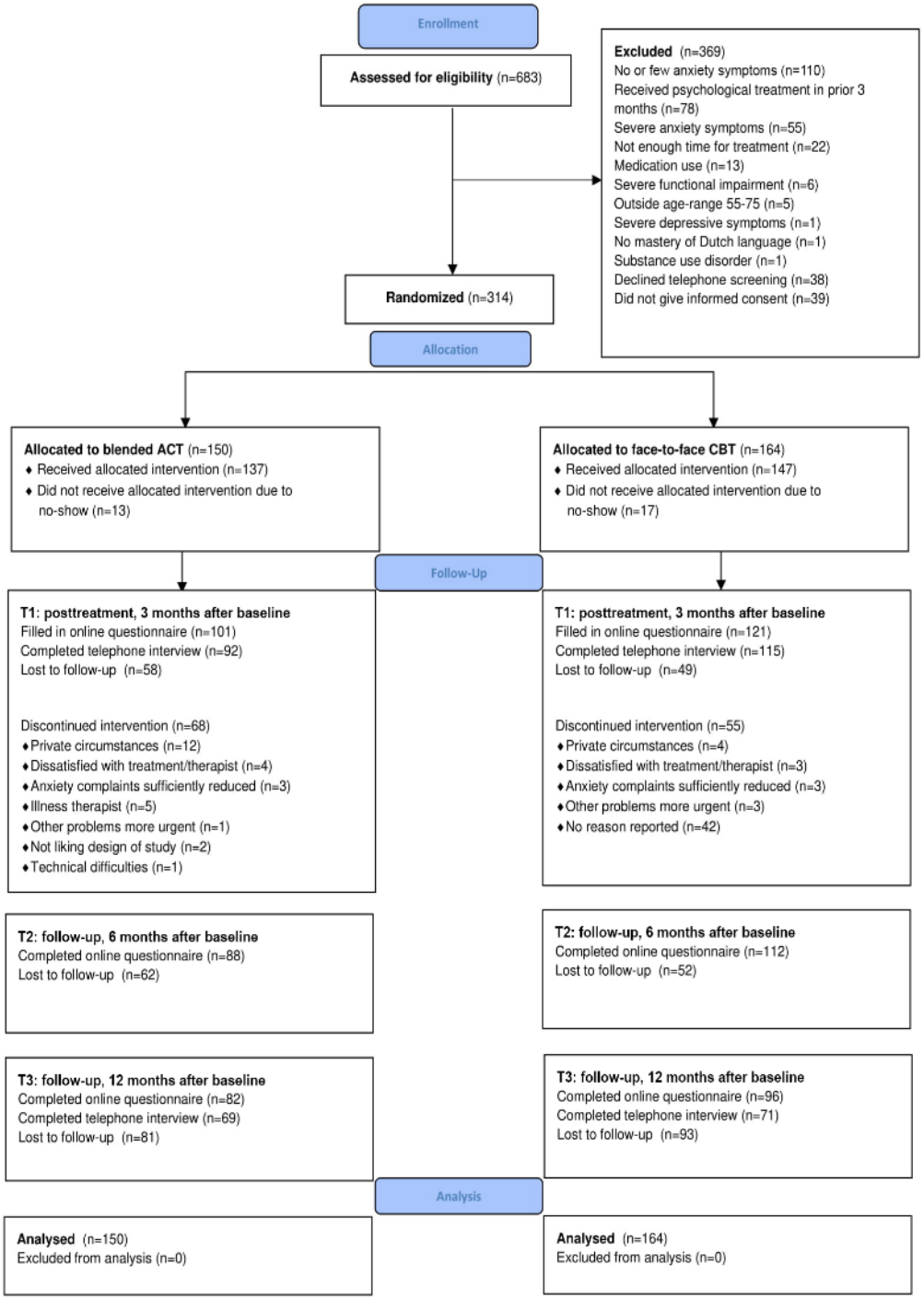


Table 2. Baseline characteristics of study sample.

Characteristics	Blended ACT[a] (n=150)	CBT[b] (n=164)	Total sample (n=314)
Age (years)			
Mean (SD)	62.75 (5.69)	63.33 (5.71)	63.06 (5.70)
Range	55-75	55-75	55-75
Sex, n (%)			
Female	100 (66.7)	92 (56.1)	192 (61.2)
Male	50 (33.3)	72 (43.9)	122 (38.9)
Nationality, n (%)			
Dutch	149 (99.3)	159 (96.9)	308 (98.0)
Dutch and other	0 (0.0)	5 (3.0)	5 (1.6)
Other	1 (0.7)	0 (0.0)	1 (0.4)
Education, n (%)			
Low	22 (14.7)	15 (9.2)	37 (11.7)
Middle	70 (44.7)	74 (45.1)	144 (45.8)
High	56 (37.3)	74 (45.1)	130 (41.4)
Unknown	2 (0.6)	1 (0.6)	3 (0.9)
Relational status, n (%)			
Married or in a romantic relationship	120 (80.0)	129 (78.6)	249 (79.3)
Not married or in a romantic relationship	30 (20.0)	35 (21.3)	65 (20.7)
Work status, n (%)			
Paid employment	77 (51.3)	76 (46.3)	153 (48.7)
Voluntary work	49 (32.6)	56 (34.1)	105 (33.4)
No work	53 (35.3)	59 (35.9)	112 (35.6)
Living situation, n (%)			
Alone	36 (24.0)	39 (23.8)	75 (23.9)
With partner	97 (64.6)	103 (62.8)	200 (63.7)
With children	11 (7.3)	13 (7.9)	24 (7.6)
With partner and children	6 (4.0)	8 (4.8)	14 (4.5)
Other	0 (0.0)	1 (0.6)	1 (0.3)
Community-dwelling	150 (100.0)	164 (100.0)	314 (100.0)
Somatic comorbidity, n (%)			
No somatic problems	29 (19.3)	32 (19.5)	61 (19.4)
One or more somatic problems	121 (80.7)	132 (80.5)	253 (80.6)
Psychomedication use, n (%)			
SSRI[c]	10 (6.7)	12 (7.3)	22 (7.0)
Benzodiazepine	19 (12.7)	15 (9.2)	34 (10.8)
No psychotropic medication	121 (80.7)	137 (83.5)	258 (82.2)
Anxiety disorder, n (%)			
Panic disorder	10 (6.7)	7 (4.3)	17 (5.4)
Agoraphobia	5 (3.3)	5 (3.0)	10 (3.2)
Social phobia	5 (3.3)	8 (4.9)	13 (4.1)
Specific phobia	10 (6.7)	8 (4.8)	18 (5.7)

Characteristics	Blended ACT[a] (n=150)	CBT[b] (n=164)	Total sample (n=314)
OCD[d]	1 (0.7)	2 (1.2)	3 (0.9)
PTSD[e]	2 (1.3)	1 (0.6)	3 (0.9)
Illness anxiety disorder	3 (2.0)	4 (2.4)	7 (2.2)
GAD[f]	17 (11.3)	18 (10.9)	35 (11.1)
Any anxiety disorder	42 (28.0)	39 (23.8)	81 (25.8)
No anxiety disorder	108 (72.0)	125 (76.2)	233 (74.2)

[a]ACT: acceptance and commitment therapy.

[b]CBT: cognitive behavioral therapy.

[c]SSRI: selective serotonin reuptake inhibitor.

[d]OCD: obsessive-compulsive disorder.

[e]PTSD: posttraumatic stress disorder.

[f]GAD: generalized anxiety disorder.

Treatment Adherence and Study Dropout

Of the 314 participants, a total of 191 (60.8%) attended all 4 face-to-face sessions and 35 (11.1%) attended 3 sessions. Significantly more participants attended 3 or 4 sessions (ie, received PP treatment) in the CBT group than in the ACT group (CBT: 126/164, 76.8%, ACT: 100/150, 66.7%, (χ^2_1=4.0, P=.045). A total of 41 participants reported their reason for dropping out of treatment (Figure 1).

The proportion of participants who completed the T1 measurement did not differ between the groups (χ^2_1=1.6, P=.21). Baseline characteristics did not differ significantly between participants who completed T1 and those who did not. Of the 222 participants who completed T1, 201 (90.5%) attended either 3 or 4 face-to-face sessions. There was no difference between the groups regarding the time participants at T1 reported to have spent on homework exercises or completing the web-based module (F_1=1.239; P=.27).

Treatment Integrity

Mental health counselors in the ACT and CBT groups completed the treatment integrity checklist for 71.1% (315/443) and 82% (424/517) of the sessions, respectively. The ACT group indicated adherence to all the prescribed elements for 80% (252/315) of the sessions. For the CBT group, this was 85.8% (364/424) of the sessions.

Primary Outcomes

Tables 3 and 4 contain the results of the mixed models and the models' estimated mean scores. Figure 2 presents the estimated mean GAD-7 scores for all measurement moments for the 2 groups. Regardless of the condition, GAD-7 scores significantly decreased from T0 to T1 (B=−3.92, P<.001), increased significantly between T1 and T2 (B=.64, P=.02), and did not change significantly from T1 to T3 (B=−.23, P=.45). The within-group effect sizes for both conditions were large for the decreases from T0 to T1 (ACT: Cohen d=0.96; CBT: Cohen d=1.09) and small to very small for T1-T2 (ACT: Cohen d=0.10; CBT: Cohen d=0.28) and T1-T3 (ACT: Cohen d=0.11; CBT: Cohen d=0.02) changes. All time-by-condition interactions were statistically insignificant, indicating that changes in anxiety symptom severity over time did not differ between the groups.

Table 3. Mixed model analyses comparing the differences between the blended acceptance and commitment therapy and cognitive behavioral therapy group over time and between-group effect sizesa.

Outcome	Unstandardized beta coefficient B	SE	*t* test (*df*)	*P* value	Cohen *d*
GAD-7[b]					
T0-T1	−3.92	0.26	−15.01 (57)	<.001	N/A[c]
T1-T2	.64	0.28	2.29 (580)	.02	N/A
T1-T3	−0.23	0.30	−0.78 (20)	.45	N/A
T0-T1[d] condition	.18	0.52	0.35 (57)	.73	0.02
T1-T2[d] condition	−0.63	0.56	−1.13 (580)	.26	0.15
T1-T3[d] condition	−0.33	0.60	−0.54 (20)	.59	0.08
MHC-SF[e]					
T0-T1	.29	0.05	4.55 (34)	<.001	N/A
T1-T2	.00	0.06	0.01 (173)	.99	N/A
T1-T3	−0.06	0.06	−0.90 (71)	.37	N/A
T0-T1[d] condition	−0.12	0.13	−0.94 (34)	.36	0.06
T1-T2[d] condition	.03	0.12	0.24 (173)	.82	0.03
T1-T3[d] condition	.27	0.13	2.13 (71)	.04	0.29
PHQ-9[f]					
T0-T1	−3.01	0.26	−11.59 (30)	<.001	N/A
T1-T2	−0.65	0.27	−2.37 (66)	.02	N/A
T1-T3	−0.69	0.33	−2.12 (42)	.04	N/A
T0-T1[d] condition	.31	0.52	0.59 (30)	.56	0.03
T1-T2[d] condition	−0.67	0.55	−1.21 (66)	.23	0.16
T1-T3[d] condition	−0.53	0.66	−0.80 (42)	.43	0.12
SDS[g] work					
T0-T1	−1.87	0.27	−6.96 (37)	<.001	N/A
T1-T3	−.18	0.31	−0.58 (45)	.57	N/A
T0-T1[d] condition	.28	0.54	0.53 (37)	.60	0.10
T1-T3[d] condition	.64	0.62	1.03 (45)	.31	0.23
SDS social life					
T0-T1	−1.78	0.26	−6.96 (37)	<.001	N/A
T1-T3	−.15	0.27	−0.55 (27)	.59	N/A
T0-T1[d] condition	−.18	0.51	−0.35 (37)	.73	0.07
T1-T3[d] condition	0.08	0.55	0.15 (27)	.88	0.03
SDS family and/or home					
T0-T1	−1.93	0.22	−8.78 (44)	<.001	N/A
T1-T3	−.17	0.26	−0.66 (73)	.51	N/A
T0-T1[d] condition	.02	0.44	0.05 (44)	.96	0.00
T1-T3[d] condition	−.38	0.51	−0.74 (73)	.46	0.11
CSQ-8[h]					

Outcome	Unstandardized beta coefficient B	SE	*t* test (*df*)	*P* value	Cohen *d*
T1 Intercept	22.83	0.35	65.20 (34)	<.001	N/A
T1 Condition	3.19	0.70	4.58 (37)	<.001	0.78
MINI-Plus[i] (for subgroup without anxiety disorder at baseline)[d]					
T1 Intercept	−3.47	0.96	−3.60[j]	<.001	N/A
T1 Condition	1.28	0.78	1.64[j]	.10	3.59
T3 Intercept	−2.38	0.47	−5.09[j]	<.001	N/A
T3 condition	.05	0.70	0.07[j]	.941	1.05
MINI-Plus (for subgroup with anxiety disorder at baseline)[d]					
T1 intercept	−1.34	0.46	−2.93[j]	.003	N/A
T1 condition	.38	0.62	0.61[j]	.54	1.46
T3 intercept	−1.39	0.79	−1.75[j]	.08	N/A
T3 condition	−1.01	1.08	−0.94[j]	.35	2.75

[a]For the MINI-Plus and CSQ-8, the condition variable was dummy coded (CBT=0, ACT=1). B is the unstandardized coefficient. T0 stands for baseline; T1 for posttreatment; T2 and T3 for 6- and 12-month follow-up, respectively.

[b]GAD-7: Generalized Anxiety Disorder-7.

[c]N/A: not applicable.

[d]The presented b-coefficients of logistic mixed model regressions are log-its and effect sizes odds ratios.

[e]MHC-SF: Mental Health Continuum-Short Form.

[f]PHQ-9: Patient Health Questionnaire-9.

[g]SDS: Sheehan Disability Scale.

[h]CSQ-8: Client Satisfaction Questionnaire-8.

[i]MINI-Plus: Mini International Neuropsychiatric Interview-Plus.

[j]Reported values are z-values.

Table 4. Mixed model estimated means for the outcomes in both groups and within-group effect sizes.

Outcome	T0 baseline (95% CI)	T1 posttreatment (95% CI)	T2 6-month follow-up (95% CI)	T3 12-month follow-up (95% CI)	ES[a] T0-T1	ES T1-T2	ES T1-T3
GAD-7[b]							
Blended ACT[c]	8.18 (7.49-8.88)	4.35 (3.59-5.12)	4.67 (3.86-5.49)	3.96 (3.09-4.83)	0.96	0.10	0.11
CBT[d]	8.78 (8.12-9.44)	4.76 (4.06-5.47)	5.72 (4.99-6.45)	4.70 (3.89-5.50)	1.09	0.28	0.02
MHC-SF[e]							
Blended ACT	2.73 (2.54-2.91)	2.96 (2.75-3.17)	2.98 (2.76-3.19)	3.04 (2.82-3.26)	0.24	0.02	0.09
CBT	2.57 (2.40-2.74)	2.92 (2.73-3.12)	2.91 (2.72-3.10)	2.73 (2.52-2.94)	0.38	0.01	0.20
PHQ-9[f]							
Blended ACT	6.99 (6.28-7.71)	4.14 (3.30-5.00)	3.16 (2.35-3.97)	3.19 (2.3-4.06)	0.70	0.26	0.27
CBT	7.92 (7.24-8.60)	4.76 (3.97-5.55)	4.44 (3.71-5.18)	4.33 (3.52-5.14)	0.75	0.08	0.12
SDS[g] work							
Blended ACT	3.52 (2.94-4.11)	1.80 (1.16-2.44)	N/A[h]	1.94 (1.17-2.71)	0.67	N/A	0.06
CBT	3.76 (3.17-4.35)	1.75 (1.17-2.34)	N/A	1.25 (0.45-2.05)	0.82	N/A	0.24
SDS social life							
Blended ACT	4.02 (3.51-4.53)	2.16 (1.57-2.74)	N/A	2.05 (1.38-2.72)	0.75	N/A	0.04
CBT	4.08 (3.59-4.56)	2.39 (1.86-2.91)	N/A	2.20 (1.53-2.86)	0.63	N/A	0.07
SDS family and/or home							
Blended ACT	3.82 (3.30-4.33)	1.90 (1.34-2.45)	N/A	1.54 (0.84-2.23)	0.76	N/A	0.16
CBT	3.79 (3.30-4.28)	1.85 (1.35-2.35)	N/A	1.87 (1.19-2.55)	0.71	N/A	0.00
MINI[i]-plus[j]							
Blended ACT	0	0.10 (0.02-0.30)	N/A	0.09 (0.03-0.21)	N/A	N/A	N/A
CBT	0	0.02 (0.00-0.17)	N/A	0.08 (0.03-0.19)	N/A	N/A	N/A
MINI-plus[k]							
Blended ACT	1	0.28 (0.14-0.46)	N/A	0.08 (0.02-0.28)	N/A	N/A	N/A
CBT	1	0.21 (0.10-0.39)	N/A	0.20 (0.05-0.54)	N/A	N/A	N/A

[a]ES: effect size.

[b]GAD-7: Generalized Anxiety Disorder-7.

[c]ACT: acceptance and commitment therapy.

[d]CBT: cognitive behavioral therapy.

[e]MHC-SF: Mental Health Continuum-Short Form.

[f]PHQ-9: Patient Health Questionnaire-9.

[g]SDS: Sheehan Disability Scale.

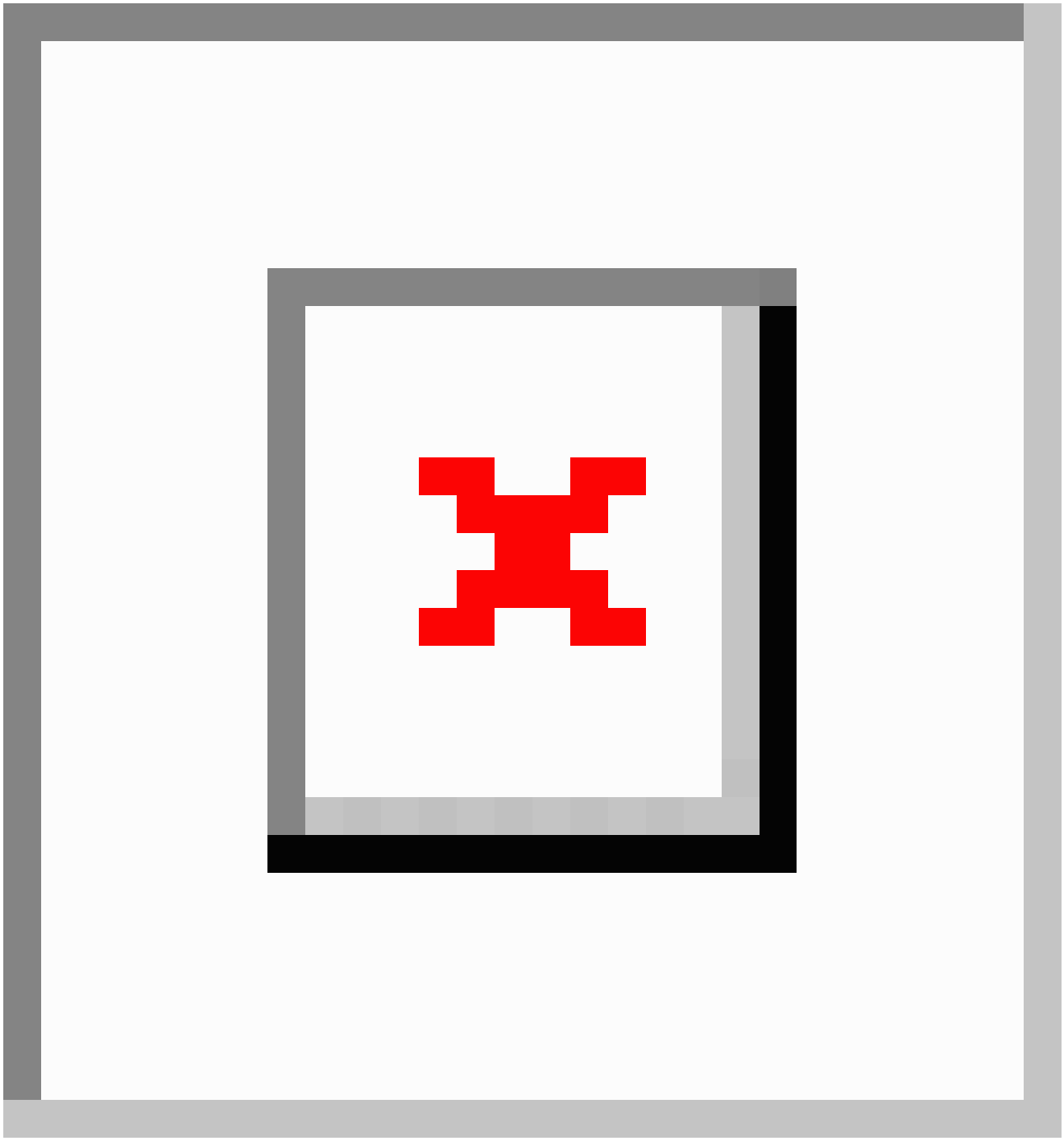
[h]N/A: not applicable.

[i]MINI: Mini-International Neuropsychiatric Interview.

[j]Probabilities of having an anxiety disorder for participants without anxiety disorder at baseline (n=233).

[k]Probabilities of having an anxiety disorder for participants with anxiety disorder at baseline (n=81).

Figure 2. Mean Generalized Anxiety Disorder-7 (GAD-7) scores at all assessments for both conditions. ACT: acceptance and commitment therapy; CBT: cognitive behavioral therapy; GAD-7: Generalized Anxiety Disorder-7.



Secondary Outcomes

The T1-T3 by condition interaction was significant for MHC-SF scores (B=.27, *P*=.03, Cohen *d*=0.29): from posttreatment to 12-month follow-up, MHC-SF scores decreased in the CBT group, whereas they increased in the ACT group. For the T0-T1 and T1-T2 intervals, no significant interactions with condition were found, but the significant main effects showed that positive mental health in both groups increased from baseline to posttreatment (B=.29, *P*<.001), and that these improvements were maintained at the month follow-up (B=.00, *P*=.99).

Time-by-condition interactions for PHQ-9 depression and SDS functional impairment were statistically insignificant. Regardless of the condition, depression severity decreased over time, as indicated by the significant main effects for all 3 time intervals (T0-T1 B=−3.01, *P*<.001; T1-T2 B=−.65, *P*=.02; T1-T3 B=−.69, *P*=.04). Functional impairment in work (B=−1.87, *P*<.001), social life (B=−1.78, *P*<.001), and family life (B=−1.93, *P*<.001) significantly decreased from baseline to posttreatment across groups. These decreases were maintained at the month follow-up (work: B=−.18, *P*=.57; social life: B=−.15, *P*=.59; family life: B=−.17, *P*=.51). In both conditions, within-group effect sizes for changes in the MHC-SF, PHQ-9, and SDS during the T0-T1 interval ranged from small to large; those for T1-T2 and T1-T3 were in the very small to small range.

Participants with anxiety disorders at baseline (n=81) had significantly higher baseline GAD-7 scores (mean 10.07, SD 4.09) than participants without an anxiety disorder (mean 1.95, SD 3.85; F_1=16.72, P<.001). Among the participants with a baseline anxiety disorder, the odds of meeting the criteria for a disorder at T1 and T3 did not differ significantly between the conditions (T1: B=.38, P=.54; T3: B=−1.01, P=.35). The odds of participants without a baseline anxiety disorder meeting the criteria for a disorder at T1 and T3 were also not significantly different in the conditions (T1 B=1.28, P=.10; T3 B=.05, P=.94).

Treatment satisfaction was significantly higher in the ACT group than in the CBT group, and the effect size of the difference was large (B=3.19, P<.001, d=0.78). No adverse events were reported.

Improvement and Clinically Significant Change

The proportions of participants with reliable anxiety symptom improvement did not differ significantly between groups (χ^2_1=0.2, P=.66). In the ACT group, 43 of the 101 (42.6%) participants showed reliable improvement at T1. In the CBT group, this was the case for 48 of the 121 (39.7%) participants. In both groups, 2 participants deteriorated. In the ACT group, 22 of the 27 (81.5%) participants with an above-cut-off GAD-7 score at baseline showed clinically significant change, whereas in the CBT group, this was the case for 27 of the 35 (77.1%) participants. These proportions did not differ significantly (χ^2_1=0.2, P=.68).

PP Analyses

PP analyses included 226 participants (ACT: n=100; CBT: n=126). PP participants did not differ significantly from other participants in terms of baseline characteristics. PP analyses replicated all the findings from the ITT analyses.

Discussion

This study evaluated the short- and long-term effectiveness of a blended ACT intervention for older adults with mild to moderately severe anxiety symptoms by comparing it with face-to-face CBT. Changes over time in anxiety symptom severity did not differ between the ACT group and CBT group. In both groups, anxiety scores significantly decreased from baseline to posttreatment, and the effect sizes for these decreases were large. At the 12-month follow-up, symptom reduction was maintained in both groups. Furthermore, rates of reliable improvement and clinically significant changes in anxiety symptoms did not differ between the groups. Analyses of secondary outcomes revealed 2 significant differences between the groups. First, improvements in positive mental health were better sustained in the long term in the ACT group. Second, treatment satisfaction was higher for the ACT intervention than for the CBT intervention. No other significant differences in secondary outcomes were found between the groups. Both groups showed significant improvements in depression severity, functional impairment, and positive mental status from baseline to posttreatment, which were mostly sustained or increased at follow-up. Finally, the proportion of participants who met the criteria for a DSM-V anxiety disorder at baseline and no longer did so after treatment did not differ between the ACT group and CBT group.

This was the first large-scale trial to evaluate an ACT intervention for anxiety in later life, and the results therefore strongly contribute to the evidence-based treatment of this highly prevalent and undertreated problem. Overall, the results show that older adults with anxiety symptoms responded similarly to the blended ACT intervention and face-to-face CBT. The insignificant differences between the ACT group and CBT group regarding the majority of outcomes add to null findings from earlier studies comparing ACT and CBT in general adult samples with anxiety symptoms or disorders [54,55]. Therefore, studies thus far have indicated that for anxious adults within a wide age range, ACT and CBT interventions are equally effective. For a more thorough understanding of the (unique) clinical value of blended ACT and face-to-face CBT for anxiety in later life, in subsequent studies we will conduct a cost-effectiveness analysis, examine their working mechanisms (mediator analyses), and determine whether they differentially affect certain subgroups of patients (moderator analyses).

A significant difference between interventions was found for positive mental health: scores from posttreatment to 1-year follow-up decreased in the CBT group and slightly increased in the ACT group. Positive mental health is an important treatment outcome, as studies have shown that after correcting for psychopathology, low levels of positive mental health are associated with more somatic diseases, increased risk of developing a mental disorder, and decreased social and work-related functioning [56]. The interaction effect found in this study is in line with the fact that stimulating people toward value-based and engaged living is an explicit goal of ACT, whereas traditional CBT is primarily focused on alleviating psychopathology [24-26]. However, assuming that ACT directly targets positive mental health, it is unexpected that there was no difference in positive mental health between the groups directly after treatment. Furthermore, the P value for the interaction was just below the α level (P=.04), and the effect size was small (d=0.29). We should, therefore, be careful not to over-interpret this finding. Therefore, the main implication of this finding is that further research into the (long-term) effects of ACT and CBT on positive mental health is warranted.

We found that treatment satisfaction was significantly higher for the blended ACT intervention than for face-to-face CBT. A pilot study on ACT for older adults with anxiety and depressive symptoms found comparable satisfaction ratings [57]. These results suggest that ACT interventions constitute a positive treatment experience for older adults, which could be related to several aspects of the treatment that have been theorized to be especially appealing to this age group [27]. However, these findings need to be interpreted with caution, as treatment satisfaction data were mainly derived from participants who attended all face-to-face sessions. As it is plausible that dropout was associated with lower treatment satisfaction and significantly more participants dropped out in the ACT group, the observed difference might, in part, be the result of selective attrition. We could not rule out this possibility because the data on reasons for dropout were incomplete.

This trial is designed to investigate the relative effectiveness of blended ACT and face-to-face CBT and does therefore not allow conclusions about the absolute effectiveness of the interventions. Still, the significant main effects of time and large within-group effect sizes for anxiety reduction from baseline to posttreatment suggest that both interventions succeeded in treating anxiety symptoms in this sample of older adults. Two earlier trials in anxious older adults found Cohen *d* values of 0.38 and 0.31 for anxiety symptom reduction (measured with the GAD-7) in waitlist conditions [12,13]. The pre-post within-group effect sizes of 0.96 (ACT) and 1.09 (CBT) in this study indicate that the symptom reduction in both conditions greatly surpassed improvements that could have been expected if participants had not received treatment.

The finding that the 2 brief, low-threshold interventions examined in this study were beneficial for a group that currently often goes untreated gives reason to be hopeful. However, to bridge the existing treatment gap, establishing the effectiveness of interventions for anxiety in later life will not suffice; efforts should also be made to increase the uptake of these interventions. In this light, it is promising that this study demonstrated a partial web-based intervention to be equally effective as face-to-face treatment, as scalable internet-based interventions might be crucial in bridging the treatment gap. As the proportion of older adults who successfully use the internet is steadily increasing [58], web-based psychological interventions seem feasible for this age group. However, it is important to note that studies have demonstrated socioeconomic disparities in internet use in older adults—higher education and income levels have been linked to more (successful) internet usage in later life [58]. This was also evident in this trial, in which internet access and basic computer skills were required to participate; more than 85% of the participants had a middle or high level of education. Large-scale implementation of internet-based psychological interventions could therefore increase health inequalities by excluding older adults without internet access or skills from treatment [59]. To improve mental health care in an inclusive manner, studies on the effectiveness and acceptability of psychological interventions for older adults with lower socioeconomic status are needed.

This study has several limitations. First, treatment integrity was assessed suboptimally because it relied on therapists' self-reports. Second, of the 35,820 people who received the information letter, only 683 registered for study participation; this is a small number considering the high prevalence of anxiety in later life [1-3]. This group is likely to differ from the study population as a whole. For example, all participants were community-dwelling, 98% were of Dutch nationality, and most had middle to high education levels. The generalizability of the findings is also limited because the more severely (psychologically and/or physically) impaired older adults and those over the age of 75 years were excluded from participation. Finally, a considerable number of participants (although comparable with other studies on internet-based and low-threshold or low-intensity interventions in general [60,61]) dropped out before completing treatment, and only one-third of them reported their reason for dropout.

In conclusion, this study is an important advancement in the evidence-based treatment of anxiety in later life. We did not find differences between blended ACT and face-to-face CBT in their effects on anxiety symptom severity and several related clinical outcomes in a large sample of older adults. In both groups, anxiety symptoms improved significantly from baseline to posttreatment, and these improvements had large effect sizes. Regarding the long-term effects on positive mental health, ACT outperformed CBT. Therefore, these findings demonstrate that blended ACT is a valuable treatment alternative to CBT for anxiety in later life, providing patients and therapists with more flexibility in deciding on the preferred intervention with regard to both treatment approach and delivery format. We will follow up this study with examinations of the cost-effectiveness, treatment mediators, and moderators of blended ACT versus CBT. Furthermore, we recommend future research to go beyond the evaluation of psychological interventions for older adults with anxiety symptoms and to focus on increasing treatment uptake in this group.

Acknowledgments

This work was supported by a grant from ZonMw (grant number 531001205).

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2226 KB - jmir_v23i3e24366_app1.pdf]

References

1. Bryant C, Jackson H, Ames D. The prevalence of anxiety in older adults: methodological issues and a review of the literature. J Affect Disord 2008 Aug;109(3):233-250. [doi: 10.1016/j.jad.2007.11.008] [Medline: 18155775]
2. Volkert J, Schulz H, Härter M, Wlodarczyk O, Andreas S. The prevalence of mental disorders in older people in Western countries - a meta-analysis. Ageing Res Rev 2013 Jan;12(1):339-353. [doi: 10.1016/j.arr.2012.09.004] [Medline: 23000171]

3. Wolitzky-Taylor KB, Castriotta N, Lenze EJ, Stanley MA, Craske MG. Anxiety disorders in older adults: a comprehensive review. Depress Anxiety 2010 Feb;27(2):190-211. [doi: 10.1002/da.20653] [Medline: 20099273]
4. de Beurs E, Beekman AT, van Balkom AJ, Deeg DJ, van Dyck R, van Tilburg W. Consequences of anxiety in older persons: its effect on disability, well-being and use of health services. Psychol Med 1999 May 1;29(3):583-593. [doi: 10.1017/s0033291799008351] [Medline: 10405079]
5. Grenier S, Préville M, Boyer R, O'Connor K, Béland S, Potvin O, et al. The impact of DSM-IV symptom and clinical significance criteria on the prevalence estimates of subthreshold and threshold anxiety in the older adult population. Am J Geriatric Psychiat 2011 Apr;19(4):316-326. [doi: 10.1097/jgp.0b013e3181ff416c]
6. Simning A, Seplaki CL. Association of the cumulative burden of late-life anxiety and depressive symptoms with functional impairment. Int J Geriatr Psychiatry 2020 Jan 28;35(1):80-90 [FREE Full text] [doi: 10.1002/gps.5221] [Medline: 31650615]
7. Lenze EJ, Wetherell JL. Bringing the bedside to the bench, and then to the community: a prospectus for intervention research in late-life anxiety disorders. Int J Geriatr Psychiatry 2009 Jan;24(1):1-14 [FREE Full text] [doi: 10.1002/gps.2074] [Medline: 18613267]
8. Voshaar RC. Lack of interventions for anxiety in older people. Br J Psychiatry 2013 Jul 2;203(1):8-9. [doi: 10.1192/bjp.bp.113.127639] [Medline: 23818533]
9. Wuthrich VM, Frei J. Barriers to treatment for older adults seeking psychological therapy. Int Psychogeriatr 2015 Mar 5;27(7):1227-1236. [doi: 10.1017/s1041610215000241]
10. Smit F, Comijs H, Schoevers R, Cuijpers P, Deeg D, Beekman A. Target groups for the prevention of late-life anxiety. Br J Psychiatry 2007 May 2;190(5):428-434. [doi: 10.1192/bjp.bp.106.023127] [Medline: 17470958]
11. Gould RL, Coulson MC, Howard RJ. Efficacy of cognitive behavioral therapy for anxiety disorders in older people: a meta-analysis and meta-regression of randomized controlled trials. J Am Geriatr Soc 2012 Feb 27;60(2):218-229. [doi: 10.1111/j.1532-5415.2011.03824.x] [Medline: 22283717]
12. Jones SL, Hadjistavropoulos HD, Soucy JN. A randomized controlled trial of guided internet-delivered cognitive behaviour therapy for older adults with generalized anxiety. J Anxiety Disord 2016 Jan;37:1-9. [doi: 10.1016/j.janxdis.2015.10.006] [Medline: 26561733]
13. Silfvernagel K, Westlinder A, Andersson S, Bergman K, Diaz Hernandez R, Fallhagen L, et al. Individually tailored internet-based cognitive behaviour therapy for older adults with anxiety and depression: a randomised controlled trial. Cogn Behav Ther 2018 Jul 25;47(4):286-300. [doi: 10.1080/16506073.2017.1388276] [Medline: 29068266]
14. Staples LG, Fogliati VJ, Dear BF, Nielssen O, Titov N. Internet-delivered treatment for older adults with anxiety and depression: implementation of the Wellbeing Plus Course in routine clinical care and comparison with research trial outcomes. BJPsych Open 2016 Sep 2;2(5):307-313 [FREE Full text] [doi: 10.1192/bjpo.bp.116.003400] [Medline: 27703794]
15. Dear BF, Zou JB, Ali S, Lorian CN, Johnston L, Sheehan J, et al. Clinical and cost-effectiveness of therapist-guided internet-delivered cognitive behavior therapy for older adults with symptoms of anxiety: a randomized controlled trial. Behav Ther 2015 Mar;46(2):206-217. [doi: 10.1016/j.beth.2014.09.007] [Medline: 25645169]
16. Trimbos-instituut. Landelijke Stuurgroep Multidisciplinaire Richtlijnontwikkeling in de GGZ Addendum ouderen bij de multidisciplinaire richtlijn Angststoornissen. Updated August 12 2008:1-86 [FREE Full text]
17. NHS England, NHS Improvement. Mental Health in Older People A Practice Primer. Updated September 12 2017:1-40 [FREE Full text]
18. Katzman MA, Bleau P, Blier P, Chokka P, Kjernisted K, Van Ameringen M. Canadian clinical practice guidelines for the management of anxiety, posttraumatic stress and obsessive-compulsive disorders. BMC Psychiatry 2014;14(Suppl 1):S1. [doi: 10.1186/1471-244x-14-s1-s1]
19. Kishita N, Laidlaw K. Cognitive behaviour therapy for generalized anxiety disorder: is CBT equally efficacious in adults of working age and older adults? Clin Psychol Rev 2017 Mar;52:124-136. [doi: 10.1016/j.cpr.2017.01.003] [Medline: 28119196]
20. Wetherell JL, Petkus AJ, Thorp SR, Stein MB, Chavira DA, Campbell-Sills L, et al. Age differences in treatment response to a collaborative care intervention for anxiety disorders. Br J Psychiatry 2013 Jul 2;203(1):65-72 [FREE Full text] [doi: 10.1192/bjp.bp.112.118547] [Medline: 23580378]
21. A-Tjak JG, Davis ML, Morina N, Powers MB, Smits JA, Emmelkamp PM. A meta-analysis of the efficacy of acceptance and commitment therapy for clinically relevant mental and physical health problems. Psychother Psychosom 2015 Dec 24;84(1):30-36. [doi: 10.1159/000365764] [Medline: 25547522]
22. Swain J, Hancock K, Hainsworth C, Bowman J. Acceptance and commitment therapy in the treatment of anxiety: a systematic review. Clin Psychol Rev 2013 Dec;33(8):965-978. [doi: 10.1016/j.cpr.2013.07.002] [Medline: 23999201]
23. Hayes SC. Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. Behavior Therapy 2004;35(4):639-665. [doi: 10.1016/s0005-7894(04)80013-3]
24. Bohlmeijer ET, Lamers SM, Fledderus M. Flourishing in people with depressive symptomatology increases with Acceptance and Commitment Therapy. Post-hoc analyses of a randomized controlled trial. Behav Res Ther 2015 Feb;65:101-106. [doi: 10.1016/j.brat.2014.12.014] [Medline: 25596344]

25. Fledderus M, Bohlmeijer ET, Smit F, Westerhof GJ. Mental health promotion as a new goal in public mental health care: a randomized controlled trial of an intervention enhancing psychological flexibility. Am J Public Health 2010 Dec;100(12):2372. [doi: 10.2105/ajph.2010.196196]
26. Keyes CL. Mental illness and/or mental health? Investigating axioms of the complete state model of health. J Consult Clin Psychol 2005 Jun;73(3):539-548. [doi: 10.1037/0022-006x.73.3.539]
27. Petkus AJ, Wetherell JL. Acceptance and commitment therapy with older adults: rationale and considerations. Cogn Behav Pract 2013 Feb;20(1):47-56 [FREE Full text] [doi: 10.1016/j.cbpra.2011.07.004] [Medline: 26997859]
28. Wetherell JL, Afari N, Ayers CR, Stoddard JA, Ruberg J, Sorrell JT, et al. Acceptance and Commitment Therapy for generalized anxiety disorder in older adults: a preliminary report. Behav Ther 2011 Mar;42(1):127-134 [FREE Full text] [doi: 10.1016/j.beth.2010.07.002] [Medline: 21292059]
29. Hohaus L, Spark J. European Psychiatry 2013 Jan;28:1. [doi: 10.1016/s0924-9338(13)77295-x]
30. Shallcross AJ, Ford BQ, Floerke VA, Mauss IB. Getting better with age: the relationship between age, acceptance, and negative affect. J Pers Soc Psychol 2013 Apr;104(4):734-749 [FREE Full text] [doi: 10.1037/a0031180] [Medline: 23276266]
31. Cheavens JS, Strunk DR, Lazarus SA, Goldstein LA. The compensation and capitalization models: a test of two approaches to individualizing the treatment of depression. Behav Res Ther 2012 Nov;50(11):699-706. [doi: 10.1016/j.brat.2012.08.002] [Medline: 22982085]
32. Wingate LR, Van Orden KA, Joiner TE, Williams FM, Rudd MD. Comparison of compensation and capitalization models when treating suicidality in young adults. J Consult Clin Psychol 2005 Aug;73(4):756-762. [doi: 10.1037/0022-006x.73.4.756]
33. Ciarrochi J, Bilich L, Godsel C. Psychological Flexibility as a Mechanism of Change in Acceptancecommitment Therapy. Baer R. Ed. Assessing Mindfulness and Acceptance: Illuminating the Processes of Change. Oakland CA: New Harbinger Publications Inc; 2010.
34. OECD. Mental Health and Work: Netherlands. Paris: OECD Publishing Paris; 2014.
35. NHG. Third revision of the NHG guideline Anxiety disorders. Updated May. 2019. URL: https://richtlijnen.nhg.org/standaarden/angst [accessed 2020-05-20]
36. Trimbos-instituut. Landelijke Stuurgroep Multidisciplinaire Richtlijnontwikkeling in de GGZ Addendum ouderen bij de multidisciplinaire richtlijn Angststoornissen. Updated August 12 2008:1-86 [FREE Full text]
37. Witlox M, Kraaij V, Garnefski N, de Waal MWM, Smit F, Hoencamp E, et al. An Internet-based Acceptance and Commitment Therapy intervention for older adults with anxiety complaints: study protocol for a cluster randomized controlled trial. Trials 2018;19(1):1-14. [doi: 10.1186/s13063-018-2731-3]
38. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2020.
39. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
40. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
41. Leon AC, Olfson M, Portera L, Farber L, Sheehan DV. Assessing psychiatric impairment in primary care with the sheehan disability scale. Int J Psychiatry Med 1997 Jul 1;27(2):93-105. [doi: 10.2190/t8em-c8yh-373n-1uwd]
42. Lecrubier Y, Sheehan D, Weiller E, Amorim P, Bonora I, Sheehan KH, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. Eur Psychiatr 2020 Apr 16;12(5):224-231. [doi: 10.1016/s0924-9338(97)83296-8]
43. Bohlmeijer, ET, Hulsbergen M. Voluit leven Mindfulness of de kunst van het aanvaarden. Amsterdam: Boom; 2008.
44. Bohlmeijer, ET, Hulsbergen M. A beginners guide to mindfulness, live in the moment. Oxford: McGraw-Hill; 2015.
45. Fledderus M, Bohlmeijer E, Pieterse M, Schreurs K. Acceptance and commitment therapy as guided self-help for psychological distress and positive mental health: A randomized controlled trial. Psychological Medicine. 2012. URL: https://doi.org/10.1017/S0033291711001206
46. Pots W, Fledderus M, Meulenbeek PA, ten Klooster PM, Schreurs KM, Bohlmeijer ET. Acceptance and commitment therapy as a web-based intervention for depressive symptoms: randomised controlled trial. Br J Psychiatry 2016 Jan 02;208(1):69-77. [doi: 10.1192/bjp.bp.114.146068] [Medline: 26250745]
47. Lamers SM, Westerhof GJ, Bohlmeijer ET, ten Klooster PM, Keyes CL. Evaluating the psychometric properties of the Mental Health Continuum-Short Form (MHC-SF). J Clin Psychol 2011 Jan 22;67(1):99-110. [doi: 10.1002/jclp.20741] [Medline: 20973032]
48. Attkisson C, Greenfield T. The UCSF Client Satisfaction ScalesThe Client Satisfaction Questionnaire-8. In: Maruish ME. ed. The use of psychological testing for treatment planning and outcomes assessment Mahwah. NJ: Lawrence Erlbaum Associates; 1999:1.
49. Buuren, van S. Flexible Imputation of Missing Data. London: CRC press; 2012.
50. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
51. Sawilowsky SS. New Effect Size Rules of Thumb. J Mod App Stat Meth 2009 Nov 1;8(2):597-599. [doi: 10.22237/jmasm/1257035100]

52. Maassen GH. The standard error in the Jacobson and Truax Reliable Change Index: the classical approach to the assessment of reliable change. J Int Neuropsychol Soc 2004 Oct 1;10(6):888-893. [doi: 10.1017/s1355617704106097] [Medline: 15637779]
53. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology 1991;59(1):12-19. [doi: 10.1037/0022-006x.59.1.12]
54. Arch JJ, Eifert GH, Davies C, Plumb Vilardaga JC, Rose RD, Craske MG. Randomized clinical trial of cognitive behavioral therapy (CBT) versus acceptance and commitment therapy (ACT) for mixed anxiety disorders. J Consult Clin Psychol 2012 Oct;80(5):750-765 [FREE Full text] [doi: 10.1037/a0028310] [Medline: 22563639]
55. Forman EM, Herbert JD, Moitra E, Yeomans PD, Geller PA. A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. Behav Modif 2007 Nov 26;31(6):772-799. [doi: 10.1177/0145445507302202] [Medline: 17932235]
56. Keyes CL. Promoting and protecting mental health as flourishing: a complementary strategy for improving national mental health. Am Psychol 2007;62(2):95-108. [doi: 10.1037/0003-066x.62.2.95]
57. Davison TE, Eppingstall B, Runci S, O'Connor DW. A pilot trial of acceptance and commitment therapy for symptoms of depression and anxiety in older adults residing in long-term care facilities. Aging Ment Health 2017 Jul 4;21(7):766-773. [doi: 10.1080/13607863.2016.1156051] [Medline: 26942691]
58. Hunsaker A, Hargittai E. A review of Internet use among older adults. New Media Society 2018 Jul 16;20(10):3937-3954. [doi: 10.1177/1461444818787348]
59. Weiss D, Rydland HT, Øversveen E, Jensen MR, Solhaug S, Krokstad S. Innovative technologies and social inequalities in health: a scoping review of the literature. PLoS One 2018 Apr 3;13(4):e0195447 [FREE Full text] [doi: 10.1371/journal.pone.0195447] [Medline: 29614114]
60. Cavanagh K. Bennett-Levy J. ed. Oxford guide to low intensity CBT interventions. Oxford: Oxford University Press; 2010.
61. Melville K, Casey L, Kavanagh D. Dropout from Internet-based treatment for psychological disorders. Br J Clin Psychol 2010;49(4):455-471. [doi: 10.1348/014466509x472138]

Abbreviations

ACT: acceptance and commitment therapy
CBT: cognitive behavioral therapy
GAD: Generalized Anxiety Disorder
GAD-7: Generalized Anxiety Disorder-7
GP: general practitioner
ITT: intention-to-treat
MHC-SF: Mental Health Continuum-Short Form
MINI-Plus: Mini-International Neuropsychiatric Interview-Plus
PHQ-9: Patient Health Questionnaire-9
PP: per-protocol
RCI: reliable change index
SDS: Sheehan Disability Scale

Edited by R Kukafka; submitted 16.09.20; peer-reviewed by M Brown, A van der Horst; comments to author 10.10.20; revised version received 23.10.20; accepted 20.01.21; published 26.03.21.

Please cite as:
Witlox M, Garnefski N, Kraaij V, de Waal MWM, Smit F, Bohlmeijer E, Spinhoven P
Blended Acceptance and Commitment Therapy Versus Face-to-face Cognitive Behavioral Therapy for Older Adults With Anxiety Symptoms in Primary Care: Pragmatic Single-blind Cluster Randomized Trial
J Med Internet Res 2021;23(3):e24366
URL: https://www.jmir.org/2021/3/e24366
doi:10.2196/24366
PMID:33769293

Original Paper

Mobile HIV Testing Through Social Networking Platforms: Comparative Study

Piao-Yi Chiou[1,2], PhD; Nai-Ying Ko[3], PhD; Chien-Yu Chien[2], BSc

[1]Department of Nursing, National Taiwan University Hospital, Taipei City, Taiwan

[2]School of Nursing, College of Medicine, National Taiwan University, Taipei City, Taiwan

[3]Department of Nursing, College of Medicine, National Cheng Kung, Tainan, Taiwan

Corresponding Author:
Piao-Yi Chiou, PhD
School of Nursing
College of Medicine, National Taiwan University
No 2-1, Xuzhou Road, Zhongzheng District
Taipei City, 100
Taiwan
Phone: 886 2 23123456 ext 88427
Email: purechiou2@gmail.com

Abstract

Background: Improving HIV screening in key populations is a crucial strategy to achieve the goal of eliminating AIDS in 2030. Social networking platforms can be used to recruit high risk-taking men who have sex with men (MSM) to promote the delivery of voluntary counseling and testing (VCT) as mobile HIV testing. Therefore, client recruitment and availability of mobile HIV testing through social networking platforms requires further evaluation.

Objective: The aim of this study is to compare the effects of targeting high risk-taking MSM and HIV case finding between two mobile HIV testing recruitment approaches: through the traditional website-based approach and through social networking platforms.

Methods: A comparative study design and propensity score matching was applied. The traditional VCT model, that is, the control group, recruited MSM through a website, and a trained research assistant visited the walk-in testing station at a gay village on Friday and Saturday nights. The social networking VCT model, the experimental group, recruited MSM from social networking platforms by periodically reloading into and conducting web-based discussions on dating apps and Facebook. The participants then referred to others in their social networks via a popular messenger app in Taiwan. The test was conducted at a designated time and place during weekdays by a trained research assistant. Across both modes of contact, before the mobile HIV testing, participants needed to provide demographic characteristics and respond to a questionnaire about HIV risk-taking behaviors.

Results: We recruited 831 MSM over 6 months, with a completion rate of 8.56% (616/7200) in the traditional VCT model and 20.71% (215/1038) in the social networking VCT model. After propensity score matching, there were 215 MSM in each group (mean age 29.97, SD 7.61 years). The social networking model was more likely to reach MSM with HIV risk-taking behaviors, that is, those seeking sexual activity through social media, having multiple sexual partners and unprotected anal intercourse, having experience of recreational drug use, and never having or not regularly having an HIV test, compared with the traditional model. HIV positive rates (incidence rate ratio 3.40, 95% CI 1.089-10.584; *P*=.03) and clinic referral rates (incidence rate ratio 0.03, 95% CI 0.001-0.585; *P*=.006) were significantly higher among those in the social networking VCT model than in the traditional VCT model.

Conclusions: Through effective recruitment strategies on social networking platforms, the social networking VCT mode can be smoothly promoted, as compared with the traditional VCT model, to target high risk-taking MSM and increase testing outcomes.

(J Med Internet Res 2021;23(3):e25031) doi:10.2196/25031

KEYWORDS

HIV testing; internet-based intervention; men who have sex with men; mobile apps; mobile phone; risk-taking; social networking; voluntary counseling and testing

Introduction

Background

Improving HIV screening in key populations is a vital strategy to reach the goal of 95-95-95 and eliminating AIDS in 2030 [1]. A study estimated that 88% of people with HIV in Taiwan were aware of their infection status in 2019 [2]. In addition to the current screening model, a more effective testing approach should be applied, such that 12% of the undiagnosed targets also undergo HIV testing.

Men who have sex with men (MSM) who had demonstrated unsafe sexual behaviors constitute a major group with HIV infection in Taiwan and globally [3,4]. To improve the uptake of voluntary counseling and testing (VCT) of MSM, mobile HIV testing, which emphasizes on the use of vehicles to deliver VCT into the community, was developed [5,6]. Mobile HIV testing employs VCT outside hospitals to target different populations and hard-to-reach groups [7-10]. To improve the HIV case findings, mobile HIV testing information should be effectively disseminated.

Usually, the client awareness of mobile HIV testing mainly results from its announcement on several platforms such as websites, radio, flyers, posters, and notifications by advocates [7-10]. Due to the high availability of smartphones, social network platforms that include mobile geosocial network apps (GSN apps) with GPS (eg, Grindr, Jack'd, and Hornet) and web-based communities (eg, Facebook [FB]) have gained a large number of web-based users and have become a popular avenue for MSM to search for health information and meet sexual partners [11-14]. High risk-taking behaviors for HIV infection, such as having multiple sexual partners, low frequency of condom use, and a high percentage of recreational drug use, have increased by using social networking platforms [15-18]. Therefore, the use of social networking platforms may also help gather mobile HIV testing information and encourage VCT among high risk-taking MSM [19,20].

In this research, HIV testing-eliciting messages and web-based discussions over instant messaging through social networking platforms were used to encourage the use of HIV VCT and self-testing among MSM [21-25]. However, there is an urgent need for more effective recruitment measures with tailored operations of social networking platforms among high risk-taking MSM. MSM users on social networking platforms could change depending on the time and place of the user who logs in. Active strategies, such as periodically reloading into GSN apps at different times and locations and setting up a fun page on popular web-based communities, could increase opportunities for web-based users to expand the information exposure of mobile HIV testing. The inability to coordinate time and location is one of the major barriers for MSM to reach the VCT [26,27]. Providing more flexible and individual HIV screening times and locations for MSM users could facilitate the accessibility of mobile HIV testing.

In summary, social networking platforms could be a direct path to target high risk-taking MSM. It is necessary to promote mobile HIV testing via social networking platforms, to deliver VCT according to the assigned time and location by MSM, and to verify the effects of HIV case finding outcomes.

Objective

The objective of this research is to compare the effects of targeting high risk-taking MSM and HIV case finding between 2 mobile HIV testing recruitment approaches: through the website and through social networking platforms.

Methods

Study Design

A comparative study design was applied, wherein two VCT models were evaluated, and purposive sampling was employed. The traditional VCT model is the most popular model in Taiwan and was therefore chosen as the control group. Specifically, VCT would be applied at a certain period of time at a testing station in the community; this implementation was announced through a website. In contrast, the social networking VCT model included scheduling through social networking platforms and providing VCT at designated times and places. The most popular social networking platforms, namely, Grindr, Hornet, Jack'd, FB, and Line (a popular mobile instant messaging app in Taiwan), were selected based on the previous year's data (MOHW106-CDC-C-114-000115). The study covered the period from May 1, 2018, to November 1, 2018, in Taipei and New Taipei City. These two cities were chosen because of their higher prevalence rates of HIV infection in Taiwan.

Participants

Inclusion criteria were self-reporting as MSM, having sexual experience, being older than 20 years, and being literate. Exclusion criteria were those who self-reported as not MSM, not having any sexual experience, and previously known to be HIV positive.

Ethical Considerations

The study was approved by the institutional review board of a medical center (18MMHISO25e). All the participants' data were processed anonymously using codes. During the HIV confirmation test process, the participants did not need to disclose any personally identifiable information to the research team to maintain their privacy. After completing the consultation, questionnaire, and testing, each participant was eligible to receive a gift worth US $3.27.

Procedure and Data Collection

Traditional VCT Model

The testing information of the traditional VCT model included free and anonymous HIV rapid testing; the testing time and location were announced every day during the study period via the public website of a municipal hospital. The traditional VCT model was conveyed to a screening station located at the entrance of a gay village between 6 PM and 10 PM every Friday and Saturday in Taipei City. The participants could actively browse the website, pass and see the outreach station, or be referred by their networks to learn about the testing information and to access the VCT. Moreover, the traditional VCT model provides walk-in services without an appointment.

A trained research assistant conducted pretest counseling and explained the research design and purpose. After obtaining written informed consent, the participants were asked to complete a questionnaire, and the pre- and posttest counseling of the rapid HIV test was performed for each participant. The participants were informed of the test results immediately after the test. For the participants with a positive test result, the research assistant accompanied them to the clinics for further confirmation via a diagnostic test. For the participants with a negative result, crucial resources about HIV prevention, such as pre-exposure prophylaxis (PrEP) and postexposure prophylaxis (PEP), were introduced, and a referral was made for them.

Social Networking VCT Model

The social networking VCT model recruited MSM from social networking platforms and provided a free and anonymous rapid HIV test at a time and place designated by the participants.

A profile heading with *mobile HIV testing* and a picture of a well-trained research assistant was set up (Figure 1) in the GSN apps: Grindr, Hornet, and Jack'd. The profile clearly introduced the program's purpose; the name of the entrusted planning unit; reservation for delivering a free, anonymous, and negotiable screening time and place; and the content of HIV testing services to enhance the audience's trust [28]. The users appear on a grid with 3 to 4 profile photos in each column in Grindr, Hornet, and Jack'd, who were within a 7-mile radius from the research assistant. The main log-in location of the research assistant was the program institution at Shipai Road, Beitou District, Taipei City. The other log-in locations changed according to the designated testing locations of the participants in Taipei and New Taipei City. Two strategies of client mobilization and recruitment were applied to the GSN apps [20,29,30]: (1) reloading once every 2 hours, a total of 4 times per day, from 10 AM to noon and 1:30 PM to 7:30 PM, on weekdays, onto each app to display our heading and profile, and allow nearby web-based users who were interested in mobile HIV testing to actively tap or send a private message to the research assistant, and (2) the research assistant provided one-on-one web-based discussions by using standard contents, including the research purpose, the privacy and rights of the participants, risk-taking behavior, and the window period of HIV infection. Individual mobile HIV testing appointments were decided after discussion.

An FB fun page (Figure 2) was created with *mobile testing in Taipei City* under the top section of the page with the mobile HIV testing poster images, and a clear description of the mobile HIV testing was provided to raise awareness and build rapport with the users on FB. Potential participants on FB who were interested in mobile HIV testing could send a private message to have a one-on-one discussion and make an appointment with the research assistant using standard content. A previous study used a similar recruitment method using FB [31].

Figure 1. The profile of the mobile HIV testing in geosocial network apps.

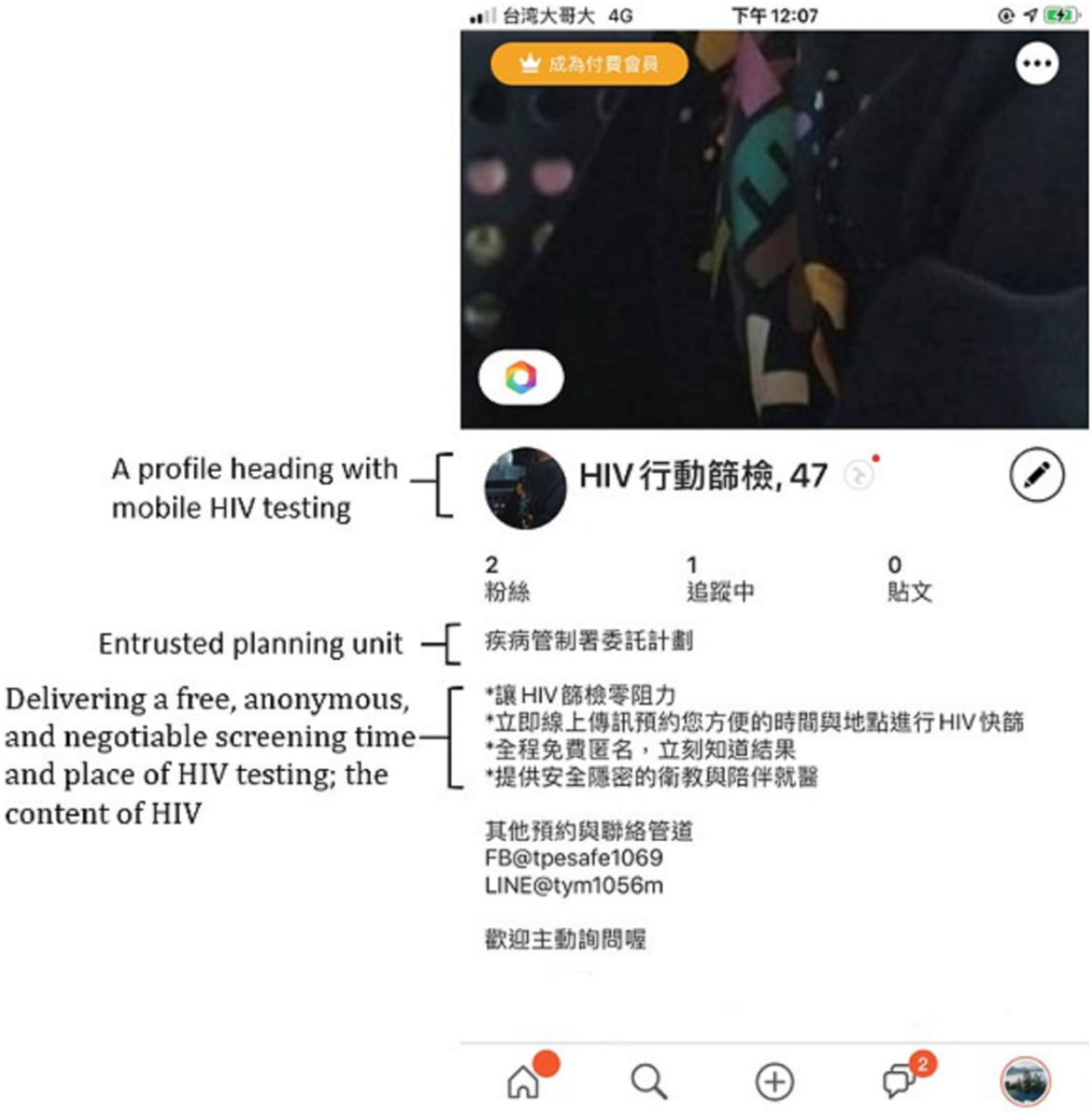


Figure 2. The profile of the mobile HIV testing on the Facebook page.

We encouraged the participants who had finished the mobile HIV testing to refer to their social network to make a mobile HIV testing appointment through a quick response code for a messenger app (Line app).

The same research assistant delivered the VCT according to the appointment at a designated time between 10 AM to noon and 1:30 PM to 7:30 PM on weekdays and at a convenient place. The ensuing test counseling, testing procedures, and posttest services were the same as in the traditional VCT model.

Table 1 shows the content comparison of the 2 VCT models.

Table 1. The content comparison of the 2 voluntary counseling and testing models.

Process	Traditional VCT[a] model	Social networking VCT model
The client mobilization and recruitment strategies	• The screening information is announced through a public website of a municipal hospital every day during the research period • Subjects pass and see the outreach station • Referral by their networks	• In GSN[b] apps: (1) reloading the user location information in the apps, namely, Grindr, Hornet, and Jack'd, once every 2 hours, a total of 4 times a day during 10 AM to noon and 1:30 PM to 7:30 PM on weekdays to display our mobile HIV testing heading and profile for nearby web-based users, and (2) providing one-on-one web-based discussions • Create a Facebook fun page and an ID in a messenger app (Line app) to provide the one-on-one private message discussion • Encourage the participants who had finished the mobile HIV testing to refer their social network through the Line app
Time of VCT	6 PM to 10 PM every Friday and Saturday	Designated by the participants during 10 AM to noon and 1:30 PM to 7:30 PM every weekday
Location of VCT	An outreach screening station at an entrance of the largest gay village in Taipei City	Designated by the participants at a convenient place in Taipei and New Taipei City
Testing appointment	No need to make an appointment	Need to make an appointment
Testing method	Rapid HIV test	Rapid HIV test
Staff	A trained research assistant	A trained research assistant
Result notification	Informed of the HIV testing result immediately after the test	Informed of the HIV testing result immediately after the test
Posttest services	The referring of confirm test and PEP[c] and PrEP[d] can be arranged directly according to the testing result	The referring of confirm test and PEP and PrEP can be arranged directly according to the testing result

[a]VCT: voluntary counseling and testing.

[b]GSN: geosocial network.

[c]PEP: postexposure prophylaxis.

[d]PrEP: pre-exposure prophylaxis.

Instrument

Demographic Characteristics Questionnaire

Participants provided demographic information, such as age, education, employment status, religion, and whether coming out of sexual orientation.

HIV Risk-Taking Behaviors Questionnaire

HIV risk-taking behaviors included days since the last unsafe sex, seeking sexual activity through social media, the nature of the relationship with the current sexual partner within the past 3 months, the frequency of anal intercourse and of condom use during anal intercourse within the past 3 months, recreational drug use, history of sexually transmitted diseases, having had a prior HIV test, regularity of this testing within the previous year, and previous PEP and PrEP.

Rapid HIV Test

DETERMINE HIV–1 and 2, antigen and antibody combo was used as the rapid screening tool in this study, which was licensed by the Taiwan Food and Drug Administration with a high sensitivity of 100% and a high specificity of 99.5% [32]. It takes only about 15 min to obtain the results, and it can be performed by trained staff.

Propensity Score Matching and Multivariable Statistical Analyses

As the subjects who accepted the anonymous HIV testing could not be randomly assigned, propensity score matching (PSM) was applied using SPSS (SPSS Inc.) software to perform the matching in a 1:1 ratio with a caliper distance of 0.001 between the 2 VCT models to control the extraneous variables and improve the homogeneity of the characteristics of the participants in the 2 VCT models. Perfect PSM was based on age, education, and employment level [33]. The comparison of demographic data and risk-taking behavior (secondary outcome) between the 2 models was analyzed using a paired *t* test for continuous variables and a chi-square test for categorical variables. Percentages were used to describe the distribution of the time and location of the social networking VCT model. All

testing locations were imported into Google Maps to calculate the coverage area of the social networking VCT model. The incidence rate ratio (IRR) and 95% CI were used to test the significance of the primary outcome of HIV testing results and the referral rates after a positive result.

Estimation of Sample Size

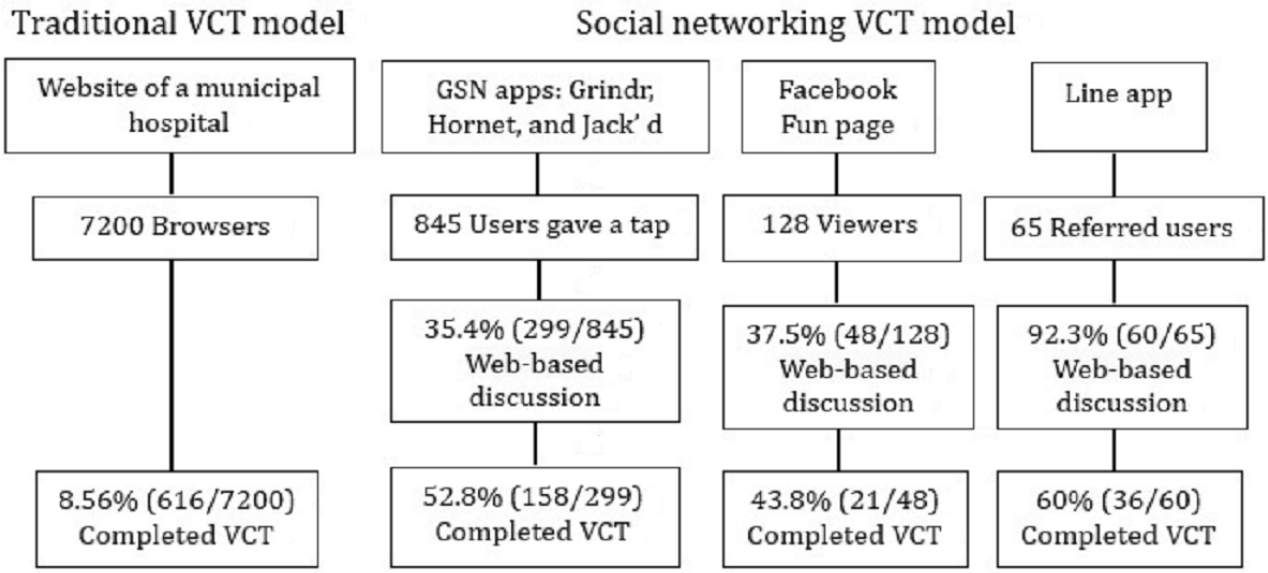
G*Power was used to estimate the matched pair sample size. On the basis of an effect size of 0.25 and power of 0.95, the sample size in each group was calculated to be 175. At a loss rate of 20%, the sample size was calculated to be at least 210 for each group.

Results

The Recruitment and Completion of 2 VCT Models

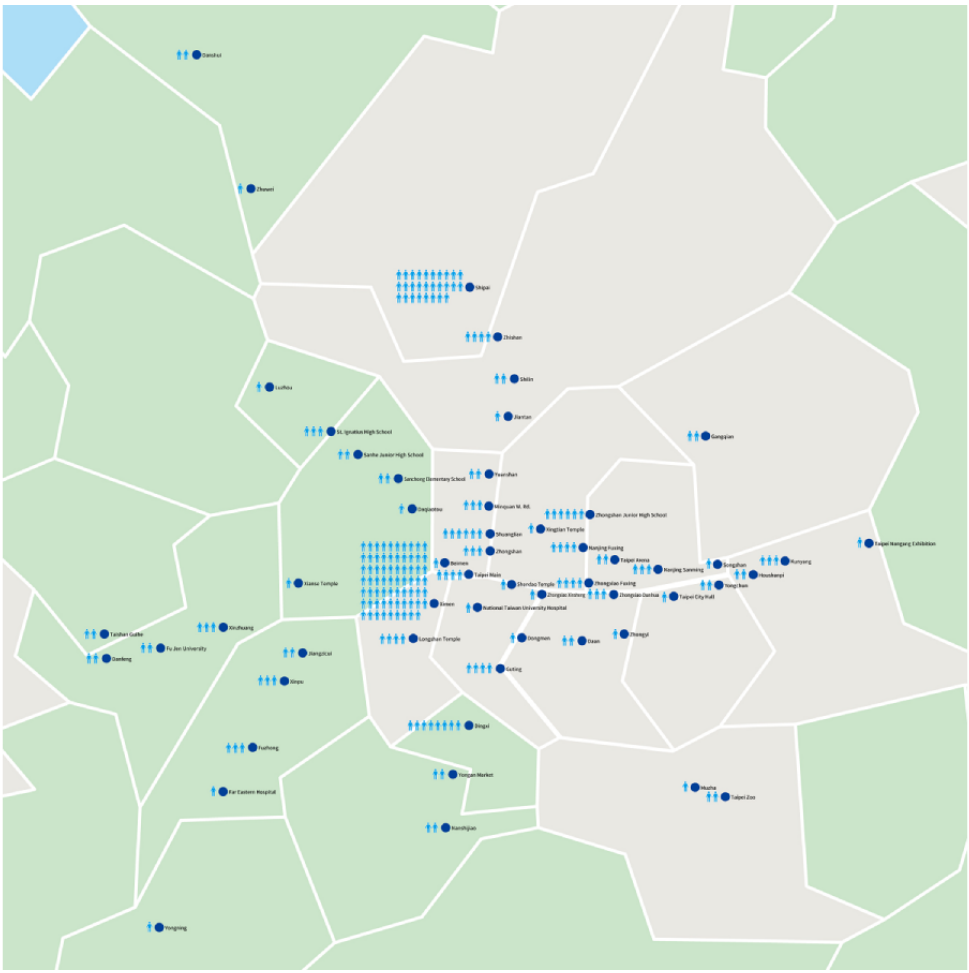
Figure 3 shows the attrition diagrams of the 2 VCT models. In total, 831 MSM participants were recruited into the 2 VCT models during the research period: 215 in the social networking VCT model and 616 in the traditional VCT model. An estimated 7200 visits, including website browsing, passing, or referrals for obtaining the VCT in the outreach station, achieved a completion rate of 8.56% (616/7200) in the traditional VCT model. Of the 215 participants in the social networking VCT model, 73.5% (158) were recruited via GSN apps, 16.7% (36) were recruited via the Line app, and 9.8% (21) were recruited via the FB fun page. A total of 845 GSN app users in Grindr, Hornet, and Jack'd actively gave us a tap, of which 35.4% (299) entered a one-on-one web-based discussion with the research assistant and 52.8% (158/299) received the VCT. After 48 private message discussions from 128 viewers through FB fun page, 44% (21/48) users received the VCT. There were 65 referred users sending the message to as through Line app, of which 92% (60) of them had one-on-one web-based discussion and 60% (36/60) of those received the VCT. The overall completion rate of the social networking VCT model, which was calculated through the number of the participants who have completed the mobile HIV testing (n=215) divided by the sum of number of the subjects who giving as the tap in GSN app (n=845), being viewers in the FB fun page (n=128), and being referred through the Line app (n=65), the result is 20.71% (215/1038). The main reasons for not accepting social networking VCT include having no response message, having been screened recently, having no risk-taking behavior, and not wanting to know the result.

Figure 3. Attrition diagram for the 2 voluntary counseling and testing models. GSN: geosocial network; VCT: voluntary counseling and testing.



The most popular designated testing period of the social networking VCT model was 1:30 PM to 5:30 PM (129/215, 60.0%), followed by 10 AM to noon (48/215, 22.3%) and 5:30 PM to 7:30 PM (38/215, 17.7%). The testing locations of the social networking VCT model and reloading locations of GSN apps by mobile phones expanded to adjacent walking areas within the 55 Taipei mass rapid transportation (MRT) stations (Figure 4). The total area of the social networking VCT model was calculated after connecting the farthest MRT station and was estimated to be 235.28 km², which covered 10.12% of the area of Taipei and New Taipei City. Figure 4 also shows the distribution of screening numbers for MSM. There were 29.3% (63/215) participants who requested mobile HIV testing of the social networking VCT model at their own house, 27.0% (58/215) at a convenience store such as 7-Eleven or Family Mart, and 24.2% (52/215) at a fast-food restaurant or café such as McDonalds or Starbucks. There were 14.0% (30/215) participants who requested mobile HIV testing in an outdoor area, such as a street side, a garden, or a school campus corner, and 5.6% (12/215) at a gay bar.

Figure 4. The distribution of tested men who have sex with men in social networking voluntary counseling and testing model in Taipei and New Taipei City.



The Comparison of the Characteristics of MSM Between the 2 VCT Models

After PSM, 215 participants with a perfect match score were selected for each model. Table 2 presents a comparison of the participants' characteristics between the 2 models. The mean age of all participants was 29.97 years (SD 7.61); most had college- or university-level education (294/430, 68.4%), were employed (296/430, 68.8%), and did not practice any religion (224/430, 52.1%). Approximately 60.5% (260/430) of the participants responded that they had experience coming out as MSM. After PSM, the demographic characteristics analysis results showed homogeneity between the 2 groups.

Table 2. The comparison of the characteristics of men who have sex with men between 2 voluntary counseling and testing models.

Variable	Total (N=430)		Traditional VCT[a] model (n=215)		Social networking VCT model (n=215)		*t* test (*df*)	Chi-square (*df*)	*P* value
	Mean (SD)	n (%)	Mean (SD)	n (%)	Mean (SD)	n (%)			
Age (years)	29.97(7.61)	N/A[b]	30.03(7.8)	N/A	29.79(7.5)	N/A	−0.39 (214)	N/A	.70
Education							N/A	0.0 (2)	>.99
Above university	N/A	86 (20.0)	N/A	43 (20.0)	N/A	43 (20.0)			
College or university	N/A	294 (68.4)	N/A	147 (68.4)	N/A	147 (68.4)			
High school or less	N/A	50 (11.6)	N/A	25 (11.6)	N/A	25 (11.6)			
Employment status							N/A	0.0 (2)	>.99
Student	N/A	92 (21.4)	N/A	46 (21.4)	N/A	46 (21.4)			
Employed	N/A	296 (68.8)	N/A	148 (68.8)	N/A	148 (68.8)			
Unemployed	N/A	42 (9.8)	N/A	21 (9.8)	N/A	21 (9.8)			
Religion							N/A	0.2 (1)	.70
Yes	N/A	206 (47.9)	N/A	101 (47.0)	N/A	105 (48.8)			
No	N/A	224 (52.1)	N/A	114 (53.0)	N/A	110 (51.2)			
Coming out of sexual orientation							N/A	0.97 (1)	.32
Yes	N/A	260 (60.5)	N/A	125 (58.1)	N/A	135 (62.8)			
No	N/A	170 (39.5)	N/A	90 (41.9)	N/A	80 (37.2)			

[a]VCT: voluntary counseling and testing.

[b]N/A: not applicable.

The Comparison of the Risk-Taking Behavior of MSM Between the 2 VCT Models

Compared with the traditional VCT model, some of the HIV risk-taking behaviors were significantly higher in the social networking VCT model (Table 3). The last unsafe sex time to date (in days) was 79.89 days shorter (t_{209}=−12.16; *P*<.001) in the social networking VCT model than in the traditional VCT model. There were 9.8% (21/215; χ^2_1=5.7; *P*=.02) more participants in the social networking VCT model who had experience seeking sex through social media as compared with the traditional VCT model. The percentage of participants having multiple relationships with current sexual partners within the past 3 months in the social networking VCT model was 17.7% (38/215) more and significantly higher than the traditional VCT model (χ^2_1=14.5; *P*=.001). The mean number of anal intercourses within the past 3 months was 1.95 times significantly higher in the social networking VCT model than in the traditional VCT model (t_{213}=3.2; *P*=.002). The social networking VCT model participants reported 10.2% (22/215) more of not using a condom every time and 3.3% (7/215) more of never used a condom during anal intercourse within the past 3 months, compared with the traditional VCT model participants (χ^2_2=8.8; *P*=.01). The experience of recreational drug use was 15.3% (33/215) more and was significantly higher among participants in the social networking VCT model than among those in the traditional VCT model (χ^2_1=16.6; *P*<.001). In total, 13.5% (29/215) more of the social networking VCT model participants reported having a history of a sexually transmitted disease, as compared with the traditional VCT model participants (χ^2_1=11.7; *P*=.001). There were 7.0% (15/215; χ^2_1=4.7; *P*=.03) and 8.8% (19/215; χ^2_1=7.1; *P*=.008) more of participants in the social networking VCT model who did not undergo HIV testing or regularly receive testing, respectively, than those in the traditional VCT model. There were no significant differences in the percentage of patients with PEP and PrEP between the two VCT models.

Table 3. The comparison of the risk-taking behavior of men who have sex with men between two voluntary counseling and testing models.

Variable	Total (N=430)		Traditional VCT[a] model (n=215)		Social networking VCT model (n=215)		*t* test (*df*)	Chi-square (*df*)	*P* value
	Mean (SD)	n (%)	Mean (SD)	n (%)	Mean (SD)	n (%)			
Days since the last unsafe sex	65.2 (76.52)	N/A[b]	107.8 (84.71)	N/A	27.91 (37.59)	N/A	−12.2 (209)	N/A	<.001
Seeking sexual activity through social media							N/A	5.7 (1)	.02
Yes	N/A	329 (76.5)	N/A	154 (71.6)	N/A	175 (81.4)			
No	N/A	101 (23.5)	N/A	61 (28.4)	N/A	40 (18.6)			
Relationship with the current sexual partner[c]							N/A	14.5 (2)	.001
Only one and stable	N/A	149 (34.7)	N/A	89 (41.1)	N/A	60 (27.9)			
Multiple	N/A	180 (41.9)	N/A	71 (33.0)	N/A	109 (50.7)			
Single	N/A	101 (23.5)	N/A	55 (25.6)	N/A	46 (21.4)			
Number of anal intercourses[c]	5.00 (6.83)	N/A	3.94 (4.24)	N/A	5.89 (8.39)	N/A	3.2 (213)	N/A	.002
Condom use frequency during anal intercourse[c]							N/A	8.8 (2)	.01
Every time used	N/A	165 (38.4)	N/A	97 (45.1)	N/A	68 (31.6)			
Not every time used	N/A	236 (54.9)	N/A	107 (49.8)	N/A	129 (60.0)			
Never used	N/A	29 (6.7)	N/A	11 (5.1)	N/A	18 (8.4)			
Experience of recreational drug use							N/A	16.6 (1)	<.001
Yes	N/A	81 (18.8)	N/A	24 (11.2)	N/A	57 (26.5)			
No	N/A	349 (81.2)	N/A	191 (88.8)	N/A	158 (73.5)			
History of sexual transmitted disease							N/A	11.7 (1)	.001
Yes	N/A	91 (21.2)	N/A	31 (14.4)	N/A	60 (27.9)			
No	N/A	339 (78.8)	N/A	184 (85.6)	N/A	155 (72.1)			
Having had a prior HIV test							N/A	4.7 (1)	.03
Yes	N/A	375 (87.2)	N/A	195 (90.7)	N/A	180 (83.7)			
No	N/A	55 (12.8)	N/A	20 (9.3)	N/A	35 (16.3)			
Having HIV test regularly (n=375)[d]							N/A	7.1 (1)	.008
Yes	N/A	268 (71.5)	N/A	151 (77.4)	N/A	117 (65.0)			
No	N/A	107 (28.5)	N/A	44 (22.6)	N/A	63 (35.0)			
Having PEP[e] previously							N/A	2.7 (1)	.10
Yes	N/A	33 (7.7)	N/A	12 (5.6)	N/A	21 (9.8)			
No	N/A	397 (92.3)	N/A	203 (94.4)	N/A	194 (90.2)			
Having PrEP[f] previously							N/A	1.8 (1)	.18
Yes	N/A	40 (9.3)	N/A	16 (7.4)	N/A	24 (11.2)			
No	N/A	390 (90.7)	N/A	199 (92.6)	N/A	191 (88.8)			

[a]VCT: voluntary counseling and testing.

[b]N/A: not applicable.

[c]Within the past 3 months.

[d]Within the past 1 year.

[e]PEP: postexposure prophylaxis.

[f]PrEP: pre-exposure prophylaxis.

Comparison of the HIV Case Finding and Clinic Referrals of the 2 VCT Models

Table 4 compares HIV testing and referrals to the clinic results between the two models. The HIV positive rate was significantly higher in the social networking VCT model (13/215, 6.0%) than in the traditional VCT model (4/215, 1.9%; IRR 3.40, 95% CI 1.089-10.584; *P*=.03). The referral rate to the clinics for confirmation diagnosis and treatment after an HIV positive result was significantly higher in the social networking VCT model (12/13, 92%) than in the traditional VCT model (1/4, 25%) (IRR 0.03, 95% CI 0.001-0.585; *P*=.006).

Table 4. The comparison of the HIV case finding and clinic referrals between the 2 voluntary counseling and testing models.

Variable	Traditional VCT[a] model (n=215), n (%)	Social networking VCT model (n=215), n (%)	IRR[b,c]	95% CI	*P* value
HIV positive	4 (1.9)	13 (6.0)	3.40	1.089-10.584	.03
Referred to the clinics for confirmation, diagnosis, and treatment	1 (25.0)	12 (92.3)	0.03	0.001-0.585	.006

[a]VCT: voluntary counseling and testing.

[b]IRR: incidence rate ratio.

[c]Incidence rate ratio was used to compare the difference of ratio.

Discussion

Principal Findings

This study applied mobile HIV testing via recruitment from social networking platforms and measured HIV case finding results. The two main findings were as follows: (1) compared with those receiving the traditional VCT model, the social networking VCT model of mobile HIV testing is more likely to reach MSM who have higher HIV risk-taking behaviors and (2) the HIV positive rates are 3 times significantly higher among those receiving the social networking VCT model than the traditional VCT model. There are three possible reasons why the HIV positive rate is higher in the social networking VCT model than in the traditional VCT model.

First, reloading into popular GSN apps in different locations could effectively enlarge the areas exposed to information about HIV testing to web-based MSM. As compared with the traditional VCT model, which provides public testing information on a certain website, the social networking VCT model was applied through the most popular GSN apps used by MSM [12,34,35]. Previous studies have indicated that MSM on social networking platforms also engage in more active unsafe sexual behaviors and do not regularly test for HIV [11-13,15-18,36]. The active presentation of heading and profiles in the social networking VCT model for nearby web-based users periodically and substantially increased mobile HIV testing information dissemination to high risk-taking MSM. Due to the delivery of VCT to the participant, the research assistant reloaded into GSN apps at an additional 55 locations, which increased the opportunity to reach broader web-based audiences [37] and avoided repeated recruitment [31]. In previous studies, pop-up and banner-paid advertisements and message sending were used to recruit MSM through GSN apps, and the range of the click-through rate was between 2.8% and 61.3% within 6 weeks to 4 months [12,22,38]. In our study, 4 times per day and a total of 1440 times of reloading into 3 GSN apps within 6 months without charge aroused 845 web-based users to click-tap, giving a click-through rate of 58.9%, which has the same effect of drawing attention as the paid advertisement. The reloading strategy of GSN apps with an attractive heading and profile of mobile HIV testing could successfully promote the attention and engagement of the research among nearby web-based MSM users, who had higher rates of HIV risk-taking behaviors.

Second, testing behavior was promoted through interactive and private message–based discussions of HIV risk-taking assessment through social networking platforms and flexible testing protocols. The HIV testing behavior of MSM could be motivated through the interactive and stimulating web-based counseling about risk-taking awareness and testing resources [23,39,40] and offering a flexible time and location for HIV screening to meet the participants' needs [41]. In our study, the average completion rate of the social networking VCT model after a one-on-one discussion with web-based users was 20.8%, which is higher than that of the traditional VCT model (8.56%). The client's own house was the most preferred place for the participants in the social networking VCT model because of privacy and convenience [42]. Although no injurious event was reported while providing the mobile HIV testing at clients' homes, it is worth noting that safety procedures were in place before going to the designated location. These included informing the team member of the time and place, setting up an emergency call button on the home page of the cell phone, and, if there were possible dangers during the test, arranging for a colleague to accompany the service provider as necessary. Compared with previous studies that applied the mobile HIV testing in a specific shelter and trunk [8-10], this study found that the mobile HIV testing could be conducted in a daily life environment, such as the seating area of convenience stores and restaurants around an MRT station in Taiwan.

Third, referrals were made via the risk-taking sexual network by those participants who had completed the mobile HIV testing of the social networking VCT model through the convenience of the Line app. MSM undergoing HIV testing usually self-identify as being exposed to risk-taking sexual behaviors and/or many sexual partners [43,44]. The acquisition of HIV testing information through network-mediated MSM could significantly predict more HIV testing behaviors than other models [45]. In addition to encouraging the sharing of HIV testing information to the network of those who had finished

the social networking VCT model, we also provided a convenient referral link by using the quick response code of the Line app. Referral of those MSM with HIV risk-taking could be facilitated while applying the information linking function and immediate web-based discussion within the Line app.

Higher referral rates to the clinic after testing HIV positive in the social networking VCT model than the traditional VCT model may be attributed to the fact that the social networking VCT model could be arranged during the daytime, providing enough time for the research assistant to accompany the positive participants immediately to the hospital for a confirmation test and further treatment. Moreover, 6.0% (12/199) of the HIV-negative participants in the social networking VCT model were referred to the PEP or PrEP clinic by our research assistant, which is higher than the traditional VCT model that had no referrals to the PEP or PrEP. The flexible HIV testing algorithms increased the chance of accessing treatment after an HIV-positive result and being given information regarding preventive measures after HIV-negative results [46].

Limitations

This study has several limitations that need to be addressed. First, the sources of the recruited participants were mainly Taipei and New Taipei City, which is a metropolitan area with convenient transportation that may limit the generalizability and extension of the results to rural areas. Second, self-report questionnaires may be skewed toward social desirability and could influence the validity of the results. Third, the implementation methods differed in several aspects between the two VCT models. These included client mobilization, recruitment methods, and testing schedules. Therefore, the effects of the social networking VCT model on HIV case findings are unclear. Fourth, the design and methods of this study require well-trained full-time personnel. Therefore, it is difficult to perform the social networking VCT model in a real-world setting when human resources and budgets are in shortage.

Conclusions

Compared with the traditional VCT model, the social networking VCT model could successfully recruit web-based MSM with a higher risk-taking of HIV by periodically reloading into social networking platforms and having discussions. The VCT is delivered in flexible testing times and locations, which increases the motivation for HIV testing behavior. Referrals to the clinic for the confirmation of diagnosis and treatment, and for PEP or PrEP, are also feasible after the social networking VCT model. The cost-effectiveness and more rigorous design of the social networking VCT model could be assessed in the future to evaluate the outcomes and increase clinical receptivity.

Acknowledgments

The authors would like to thank the Centers for Disease Control, Taiwan, ROC (MOHW107–Cover Page CDC-C-114-000106) for providing research funding. The authors are grateful to the gay bars that provide a place for the traditional VCT model. The authors are also grateful to the municipal hospital that provided HIV testing resources.

Conflicts of Interest

There are no competing financial interests in this research. The purpose of the research did not reflect the official policy or position of the organization or government.

References

1. Understanding fast-track: Accelerating action to end the AIDS epidemic by 2030. Joint United Nations Programme on HIV and AIDS. 2021. URL: https://www.unaids.org/sites/default/files/media_asset/201506_JC2743_Understanding_FastTrack_en.pdf
2. The 2020 HIV Self-Screening Program is launched! Regular screening for peace of mind. Centers for Disease Control, Taiwan. 2020 May 01. URL: https://www.cdc.gov.tw/Bulletin/Detail/P7tQd645gMwWVa8jCDkJ_Q?typeid=9 [accessed 2020-06-08]
3. AIDS statistics data. Centers for Disease Control, Taiwan. 2021 Jan 31. URL: https://www.cdc.gov.tw/Category/Page/rCV9N1rGUz9wNr8lggsh2Q [accessed 2019-09-21]
4. Joint UNPOHIVAIDS. 2020. URL: https://www.unaids.org/en/resources/fact-sheet [accessed 2020-10-21]
5. Bowles KE, Clark HA, Tai E, Sullivan PS, Song B, Tsang J, et al. Implementing rapid HIV testing in outreach and community settings: results from an advancing HIV prevention demonstration project conducted in seven U.S. cities. Public Health Rep 2008;123 Suppl 3(1):78-85 [FREE Full text] [doi: 10.1177/00333549081230S310] [Medline: 19172705]
6. Scognamiglio P, Chiaradia G, Giovanetti M, Albertini E, Camposeragna A, Farinella M, et al. HIV rapid testing in community and outreach sites: results of a nationwide demonstration project in Italy. BMC Public Health 2018 Jun 18;18(1):748 [FREE Full text] [doi: 10.1186/s12889-018-5680-6] [Medline: 29914449]
7. Bassett IV, Regan S, Luthuli P, Mbonambi H, Bearnot B, Pendleton A, et al. Linkage to care following community-based mobile HIV testing compared with clinic-based testing in Umlazi Township, Durban, South Africa. HIV Med 2014 Jul;15(6):367-372 [FREE Full text] [doi: 10.1111/hiv.12115] [Medline: 24251725]

8. Grabbe KL, Menzies N, Taegtmeyer M, Emukule G, Angala P, Mwega I, et al. Increasing access to HIV counseling and testing through mobile services in Kenya: strategies, utilization, and cost-effectiveness. J Acquir Immune Defic Syndr 2010 Jul;54(3):317-323 [FREE Full text] [doi: 10.1097/QAI.0b013e3181ced126] [Medline: 20453819]
9. Mabuto T, Latka MH, Kuwane B, Churchyard GJ, Charalambous S, Hoffmann CJ. Four models of HIV counseling and testing: utilization and test results in South Africa. PLoS One 2014;9(7):e102267 [FREE Full text] [doi: 10.1371/journal.pone.0102267] [Medline: 25013938]
10. Meehan S, Naidoo P, Claassens MM, Lombard C, Beyers N. Characteristics of clients who access mobile compared to clinic HIV counselling and testing services: a matched study from Cape Town, South Africa. BMC Health Serv Res 2014 Dec 20;14:658 [FREE Full text] [doi: 10.1186/s12913-014-0658-2] [Medline: 25526815]
11. Beymer MR, Rossi AD, Shu SB. Assessing Self-Control and Geosocial Networking App Behavior Among an Online Sample of Men Who Have Sex with Men. J Urban Health 2016 Aug;93(4):698-708 [FREE Full text] [doi: 10.1007/s11524-016-0056-7] [Medline: 27184573]
12. Holloway IW, Rice E, Gibbs J, Winetrobe H, Dunlap S, Rhoades H. Acceptability of smartphone application-based HIV prevention among young men who have sex with men. AIDS Behav 2014 Feb;18(2):285-296 [FREE Full text] [doi: 10.1007/s10461-013-0671-1] [Medline: 24292281]
13. Ko NY, Tseng PC, Huang YC, Chen YC, Hsu ST. Seeking sex partners through the internet and mobile phone applications among men who have sex with men in Taiwan. AIDS Care 2016 Jul;28(7):927-931. [doi: 10.1080/09540121.2015.1131969] [Medline: 26754350]
14. Chan PA, Towey C, Poceta J, Rose J, Bertrand T, Kantor R, et al. Online hookup sites for meeting sexual partners among men who have sex with men in Rhode Island, 2013: a call for public health action. Public Health Rep 2016;131(2):264-271 [FREE Full text] [doi: 10.1177/003335491613100210] [Medline: 26957661]
15. Landovitz RJ, Tseng CH, Weissman M, Haymer M, Mendenhall B, Rogers K, et al. Epidemiology, sexual risk behavior, and HIV prevention practices of men who have sex with men using GRINDR in Los Angeles, California. J Urban Health 2013 Aug;90(4):729-739 [FREE Full text] [doi: 10.1007/s11524-012-9766-7] [Medline: 22983721]
16. Lehmiller JJ, Ioerger M. Social networking smartphone applications and sexual health outcomes among men who have sex with men. PLoS One 2014;9(1):e86603 [FREE Full text] [doi: 10.1371/journal.pone.0086603] [Medline: 24466166]
17. O'Connor L, O'Donnell K, Barrett P, Hickson FCI, McCartney D, Quinlan M, et al. Use of geosocial networking applications is independently associated with diagnosis of STI among men who have sex with men testing for STIs: findings from the cross-sectional MSM Internet Survey Ireland (MISI) 2015. Sex Transm Infect 2019 Jun;95(4):279-284 [FREE Full text] [doi: 10.1136/sextrans-2018-053637] [Medline: 30518621]
18. Chow EP, Cornelisse VJ, Read TR, Hocking JS, Walker S, Chen MY, et al. Risk Practices in the Era of Smartphone Apps for Meeting Partners: A Cross-Sectional Study Among Men Who Have Sex with Men in Melbourne, Australia. AIDS Patient Care STDS 2016 Apr;30(4):151-154. [doi: 10.1089/apc.2015.0344] [Medline: 27028181]
19. Rendina HJ, Jimenez RH, Grov C, Ventuneac A, Parsons JT. Patterns of lifetime and recent HIV testing among men who have sex with men in New York City who use Grindr. AIDS Behav 2014 Jan;18(1):41-49 [FREE Full text] [doi: 10.1007/s10461-013-0573-2] [Medline: 23925515]
20. Sun CJ, Sutfin E, Bachmann LH, Stowers J, Rhodes SD. Comparing men who have sex with men and transgender women who use Grindr, other similar social and sexual networking apps, or no social and sexual networking apps: Implications for recruitment and health promotion. J AIDS Clin Res 2018;9(2) [FREE Full text] [doi: 10.4172/2155-6113.1000757] [Medline: 29593933]
21. Young SD, Cumberland WG, Nianogo R, Menacho LA, Galea JT, Coates T. The HOPE social media intervention for global HIV prevention in Peru: a cluster randomised controlled trial. Lancet HIV 2015 Jan;2(1):e27-e32 [FREE Full text] [doi: 10.1016/S2352-3018(14)00006-X] [Medline: 26236767]
22. Huang E, Marlin RW, Young SD, Medline A, Klausner JD. Using grindr, a smartphone social-networking application, to increase HIV self-testing among black and latino men who have sex with men in Los Angeles, 2014. AIDS Educ Prev 2016 Aug;28(4):341-350. [doi: 10.1521/aeap.2016.28.4.341] [Medline: 27427928]
23. Ko NY, Hsieh CH, Wang MC, Lee C, Chen CL, Chung AC, et al. Effects of Internet popular opinion leaders (iPOL) among Internet-using men who have sex with men. J Med Internet Res 2013;15(2):e40 [FREE Full text] [doi: 10.2196/jmir.2264] [Medline: 23439583]
24. Roberts N, Holden J, Duck T, Kitchener S. Health promotion 'on steroids': the value of an experiential approach to promote rapid HIV testing in NSW, Australia. Public Health Res Pract 2015 Mar 30;25(2):e2521522 [FREE Full text] [doi: 10.17061/phrp2521522] [Medline: 25848740]
25. Tang W, Wei C, Cao B, Wu D, Li KT, Lu H, et al. Crowdsourcing to expand HIV testing among men who have sex with men in China: A closed cohort stepped wedge cluster randomized controlled trial. PLoS Med 2018 Aug;15(8):e1002645 [FREE Full text] [doi: 10.1371/journal.pmed.1002645] [Medline: 30153265]
26. Chamie G, Clark TD, Kabami J, Kadede K, Ssemmondo E, Steinfeld R, et al. A hybrid mobile approach for population-wide HIV testing in rural east Africa: an observational study. Lancet HIV 2016 Mar;3(3):e111-e119 [FREE Full text] [doi: 10.1016/S2352-3018(15)00251-9] [Medline: 26939734]

27. Han L, Wei C, Muessig KE, Bien CH, Meng G, Emch ME, et al. HIV test uptake among MSM in China: Implications for enhanced HIV test promotion campaigns among key populations. Glob Public Health 2017 Jan;12(1):31-44 [FREE Full text] [doi: 10.1080/17441692.2015.1134612] [Medline: 26785328]
28. Gelinas L, Pierce R, Winkler S, Cohen IG, Lynch HF, Bierer BE. Using social media as a research recruitment tool: ethical issues and recommendations. Am J Bioeth 2017 Mar;17(3):3-14 [FREE Full text] [doi: 10.1080/15265161.2016.1276644] [Medline: 28207365]
29. Holloway IW. Substance use homophily among geosocial networking application using gay, bisexual, and other men who have sex with men. Arch Sex Behav 2015 Oct;44(7):1799-1811 [FREE Full text] [doi: 10.1007/s10508-015-0581-6] [Medline: 26216146]
30. Arigo D, Pagoto S, Carter-Harris L, Lillie SE, Nebeker C. Using social media for health research: methodological and ethical considerations for recruitment and intervention delivery. Digit Health 2018;4:2055207618771757 [FREE Full text] [doi: 10.1177/2055207618771757] [Medline: 29942634]
31. Iribarren SJ, Ghazzawi A, Sheinfil AZ, Frasca T, Brown W, Lopez-Rios J, et al. Mixed-method evaluation of social media-based tools and traditional strategies to recruit high-risk and hard-to-reach populations into an HIV prevention intervention study. AIDS Behav 2018 Jan;22(1):347-357 [FREE Full text] [doi: 10.1007/s10461-017-1956-6] [Medline: 29124420]
32. Determine HIV 1/2 Ag/Ab Combo. Food and Drug Administration, Ministry Health and Welfare, Taiwan. 2015. URL: https://info.fda.gov.tw/MLMS/H0001D.aspx?Type=Lic&LicId=06015099 [accessed 2020-06-08]
33. Shi L, Kanouse D, Baldwin S, Kim J. Perceptions of HIV/AIDS in one's community predict HIV testing. AIDS Behav 2012 Oct;16(7):1926-1933 [FREE Full text] [doi: 10.1007/s10461-012-0243-9] [Medline: 22767029]
34. Martinez O, Wu E, Shultz AZ, Capote J, López RJ, Sandfort T, et al. Still a hard-to-reach population? Using social media to recruit Latino gay couples for an HIV intervention adaptation study. J Med Internet Res 2014 Apr;16(4):e113 [FREE Full text] [doi: 10.2196/jmir.3311] [Medline: 24763130]
35. Baral S, Turner RM, Lyons CE, Howell S, Honermann B, Garner A, et al. Population size estimation of gay and bisexual men and other men who have sex with men using social media-based platforms. JMIR Public Health Surveill 2018 Feb 08;4(1):e15 [FREE Full text] [doi: 10.2196/publichealth.9321] [Medline: 29422452]
36. Choi EP, Wong JY, Fong DY. The use of social networking applications of smartphone and associated sexual risks in lesbian, gay, bisexual, and transgender populations: a systematic review. AIDS Care 2017 Feb;29(2):145-155. [doi: 10.1080/09540121.2016.1211606] [Medline: 27454158]
37. Delaney KP, Kramer MR, Waller LA, Flanders WD, Sullivan PS. Using a geolocation social networking application to calculate the population density of sex-seeking gay men for research and prevention services. J Med Internet Res 2014;16(11):e249 [FREE Full text] [doi: 10.2196/jmir.3523] [Medline: 25406722]
38. Macapagal K, Coventry R, Puckett JA, Phillips G, Mustanski B. Geosocial networking app use among men who have sex with men in serious romantic relationships. Arch Sex Behav 2016 Aug;45(6):1513-1524 [FREE Full text] [doi: 10.1007/s10508-016-0698-2] [Medline: 26969319]
39. Krueger EA, Chiu CJ, Menacho LA, Young SD. HIV testing among social media-using Peruvian men who have sex with men: correlates and social context. AIDS Care 2016 Oct;28(10):1301-1305 [FREE Full text] [doi: 10.1080/09540121.2016.1178699] [Medline: 27140820]
40. Choi K, Lui H, Guo Y, Han L, Mandel JS. Lack of HIV testing and awareness of HIV infection among men who have sex with men, Beijing, China. AIDS Educ Prev 2006 Feb;18(1):33-43. [doi: 10.1521/aeap.2006.18.1.33] [Medline: 16539574]
41. Vu BN, Green KE, Thi Thu Phan H, Hung Tran M, Van Ngo H, Hai Vo S, et al. Lay provider HIV testing: a promising strategy to reach the undiagnosed key populations in Vietnam. PLoS One 2018;13(12):e0210063 [FREE Full text] [doi: 10.1371/journal.pone.0210063] [Medline: 30596777]
42. Miners A, Nadarzynski T, Witzel C, Phillips AN, Cambiano V, Rodger AJ, et al. Preferences for HIV testing services among men who have sex with men in the UK: A discrete choice experiment. PLoS Med 2019 Apr;16(4):e1002779 [FREE Full text] [doi: 10.1371/journal.pmed.1002779] [Medline: 30973868]
43. LeViere A, Donovan J, Wilkin A, Keller J, Parnell H, Sampson L, et al. Results of a Social Network Testing Intervention for HIV in Infectious Disease Clinics. AIDS Behav 2019 Jan;23(Suppl 1):48-51 [FREE Full text] [doi: 10.1007/s10461-018-2178-2] [Medline: 29872998]
44. Yamanis TJ, Dervisevic E, Mulawa M, Conserve DF, Barrington C, Kajula LJ, et al. Social Network Influence on HIV Testing Among Urban Men in Tanzania. AIDS Behav 2017 Apr;21(4):1171-1182 [FREE Full text] [doi: 10.1007/s10461-016-1513-8] [Medline: 27506817]
45. Veinot TC, Caldwell E, Loveluck J, Arnold MP, Bauermeister J. HIV Testing Behavior and Social Network Characteristics and Functions Among Young Men Who have Sex with Men (YMSM) in Metropolitan Detroit. AIDS Behav 2016 Nov;20(11):2739-2761 [FREE Full text] [doi: 10.1007/s10461-016-1296-y] [Medline: 26837634]
46. Fuqua V, Chen Y, Packer T, Dowling T, Ick TO, Nguyen B, et al. Using social networks to reach Black MSM for HIV testing and linkage to care. AIDS Behav 2012 Feb;16(2):256-265 [FREE Full text] [doi: 10.1007/s10461-011-9918-x] [Medline: 21390535]

Abbreviations

FB: Facebook
GSN: geosocial network
IRR: incidence rate ratio
MRT: mass rapid transportation
MSM: men who have sex with men
PEP: postexposure prophylaxis
PrEP: pre-exposure prophylaxis
PSM: propensity score matching
VCT: voluntary counseling and testing

Edited by R Kukafka; submitted 15.10.20; peer-reviewed by YH Lai, R Giguere; comments to author 31.12.20; revised version received 20.01.21; accepted 31.01.21; published 26.03.21.

Please cite as:
Chiou PY, Ko NY, Chien CY
Mobile HIV Testing Through Social Networking Platforms: Comparative Study
J Med Internet Res 2021;23(3):e25031
URL: https://www.jmir.org/2021/3/e25031
doi:10.2196/25031
PMID:33769298

Original Paper

A Smartphone-Delivered Ecological Momentary Intervention for Problem Gambling (GamblingLess: Curb Your Urge): Single-Arm Acceptability and Feasibility Trial

Chloe O Hawker[1], BA (Hons); Stephanie S Merkouris[1], BPsychSc, GradDipPsych, PhD; George J Youssef[1,2], BASc, MBiostats, MPsych, PhD; Nicki A Dowling[1,3], BSc, GradDipAppChildPsych, PhD

[1]Deakin University, Geelong, Australia

[2]Royal Children's Hospital, Centre for Adolescent Health, Murdoch Children's Research Institute, Melbourne, Australia

[3]Melbourne Graduate School of Education, University of Melbourne, Parkville, Australia

Corresponding Author:
Chloe O Hawker, BA (Hons)
Deakin University
1 Gheringhap St
Geelong, 3220
Australia
Phone: 61 03 9244 5610
Email: chawker@deakin.edu.au

Abstract

Background: Low uptake rates of traditional gambling treatments highlight the need for innovative treatment modalities. Smartphone apps can provide unprecedented access to real-time ecological momentary interventions (EMIs) delivered in people's everyday lives.

Objective: This study aims to examine the acceptability, feasibility, and preliminary effectiveness of GamblingLess: Curb Your Urge, the first smartphone app–delivered EMI that aims to prevent gambling episodes by reducing craving intensity in people seeking help for gambling problems.

Methods: This study was a single-arm, 5-week acceptability and feasibility trial (1-week baseline and 4-week intervention periods) involving ecological momentary assessments (EMAs) delivered 3 times daily. The EMAs measured gambling episodes, cravings, and self-efficacy. Web-based evaluations at baseline, postintervention, and 1-month follow-up measured gambling outcomes (severity, cravings, frequency, expenditure, and self-efficacy) and the intervention's perceived helpfulness, relevance, burden, satisfaction, and impact in relation to gambling cravings.

Results: A total of 36 participants, of whom 22/36 (61%) were male and 34/36 (94%) were problem gamblers, completed the baseline measures, with 61% (22/36) completing the postintervention evaluation and 58% (21/36) completing the follow-up evaluation. The intervention was considered acceptable, as participants perceived all intervention content to be above average in helpfulness and the EMA to be highly relevant but somewhat burdensome. Participants reported that they were satisfied with the intervention and that the intervention improved their knowledge, attitudes, awareness, behavior change, intention to change, and help-seeking behavior for gambling cravings. Regarding the intervention's feasibility, compliance rates for the EMA (51%) and EMI (15%) were low; however, the intervention was used 166 times, including 59 uses within 60 minutes of EMA completion and 107 on-demand uses. Regarding the intervention's preliminary effectiveness, descriptive EMA data showed that, compared with the baseline period, 71% and 72% reductions in the average number of gambling episodes and craving occurrences were reported in the intervention period, respectively. In addition, clustered paired-sample two-tailed *t* tests revealed a significant 5.4% reduction in real-time craving intensity (*P*=.01) immediately after intervention use, which increased to 10.5% (*P*=.01), where use was recommended based on craving occurrence. At the group level, significant medium-to-large reductions were observed in mean gambling symptom severity (*P*=.01 and .003), cravings (*P*=.03 and .02), frequency (*P*=.01 and .004), and expenditure (*P*=.04 and .003) at postintervention and follow-up; moreover, increased mean gambling self-efficacy and craving self-efficacy (*P*=.01 and .01) were observed at postintervention and increased gambling self-efficacy (*P*=.04) was observed at follow-up. At the individual level, over a quarter of participants (6/22, 27% to 10/21, 48%) could be categorized as *recovered* or *improved* regarding their gambling symptom severity and cravings.

Conclusions: The results support the acceptability, feasibility, and preliminary effectiveness of this app-delivered EMI for preventing gambling episodes through craving management in people with gambling problems, which has implications for extending the reach of evidence-based treatment to moments of vulnerability in people's everyday lives.

(J Med Internet Res 2021;23(3):e25786) doi:10.2196/25786

KEYWORDS

gambling; craving; urge; self-efficacy; relapse; smartphone; self-help; treatment; ecological momentary assessment; ecological momentary intervention; mobile phone

Introduction

Background

Problem gambling is characterized by persistent and recurrent difficulties in limiting time and/or money spent gambling, which impacts the individual, their friends and family, and the community [1]. The harms associated with problem gambling are widespread, including financial difficulties, relationship conflict, emotional distress, health decrements, reduced productivity and job losses, and criminal activities [2]. Standardized prevalence rates estimate that 2.3% of the global population display past-year problem gambling. Australian rates estimate 0.4% to 0.6% for problem gambling, 1.9% to 3.7% for moderate-risk gambling, and 3% to 7.7% for low-risk gambling [3,4]. As such, problem gambling is a major public health issue with a burden of harm on par with depression and alcohol use disorders and almost twice that of substance use disorders, bipolar disorder, eating disorders, and schizophrenia combined [5].

Treatment of Problem Gambling

Despite meta-analytic evidence indicating the efficacy of face-to-face delivered cognitive behavioral therapy (CBT) and motivational interviewing (MI) in gambling treatment [6-8], few people access face-to-face services (8%-16%) [9]. Commonly reported service barriers include personal reasons, such as shame, stigma, and a desire for self-management, and resource, geographic, and time constraints [9-12]. To overcome these barriers, novel treatment approaches that extend the reach of evidence-based treatment beyond the confines of face-to-face services are receiving increased attention. To date, such research has focused on web-based self-directed interventions, with a recent systematic review [13] demonstrating that the 2 available high-intensity, structured web-based interventions for problem gambling [14,15] were as effective as face-to-face services for reducing problem gambling severity, gambling frequency, and financial loss at posttreatment.

Smartphone App–Delivered Ecological Momentary Interventions

Smartphone apps are a particularly advantageous modality for delivering self-directed interventions because of their ubiquitous use and extensive global networks that offer unprecedented treatment access [16,17]. As such, apps can serve as a "conduit for intervention any time and in almost any location" [18], which is considered important for preventing gambling behavior and relapse [19,20]. There is growing evidence for the acceptability, feasibility, and preliminary effectiveness of smartphone-delivered interventions for numerous chronic medical conditions [21-23] and mental health concerns [24,25]. Despite its promise, only a small number of apps delivering adjunctive [26,27] or stand-alone [28-30] interventions have been developed for problem gambling. Only 2 of these apps have undergone initial testing [26,27], with the results supporting their utility for increasing therapeutic homework completion [26] and reducing problem gambling severity [27]. Although these findings demonstrate the promise of app-delivered gambling interventions, they deliver static content that does not embrace the full potential of apps for delivering dynamic interventions when and where people need them.

Smartphone apps can deliver dynamic real-time interventions, termed *ecological momentary interventions* (EMIs), that aim to provide the right type and amount of in-the-moment support to people in their everyday lives [31]. To identify when an EMI is needed, apps can administer real-time *ecological momentary assessments* (EMAs) of relevant constructs at random or prespecified times. On the basis of the EMA results, apps can apply decision rules to determine the type, timing, and intensity of EMI delivered [31]. Smartphone app–delivered EMIs are increasingly employed in the broader mental health and addiction fields (eg, depression and anxiety [32], psychotic disorders [33], and alcohol and substance use [34-38]). Promisingly, addiction studies have demonstrated their preliminary effectiveness in identifying high-risk situations, reducing use and relapse incidence, and practicing real-time coping [34-37].

To the authors' knowledge, however, there are only 2 smartphone app–delivered EMIs in the gambling field [39,40]. The first app, *Jeu-contrôle*, aims to support people to adhere to their gambling time and money limits by delivering personalized feedback about current gambling behavior compared with preset limits recorded during EMAs [39]; however, this app has not yet been evaluated. In contrast, the research team has developed and completed usability testing of GamblingLess: Curb Your Urge [40], the first smartphone-delivered EMI, tailored by responses to prompted EMA, for gambling craving management.

GamblingLess: Curb Your Urge

GamblingLess: Curb Your Urge was developed as an app-delivered intervention within a suite of evidence-based web and mobile CBT and MI programs (GamblingLess) for problem gambling [41]. The intervention was designed to specifically target gambling cravings based on the relapse prevention model [42] and recent evidence that real-time cravings and self-efficacy to resist cravings were the strongest predictors of subsequent gambling behavior [43,44]. Ultimately, the intervention aims to prevent cravings from transitioning into gambling episodes

by identifying when a consumer is experiencing a craving (via EMA) and subsequently providing strategies to manage their craving in the moment (via EMI). Similar to other app-delivered EMIs [18,34], the intervention provides a hybrid model, comprising a *push* intervention, whereby the app recommends an intervention at times of identified need, and a *pull* intervention, whereby consumers can access an intervention *on demand*, whenever and wherever they need it, without needing to complete an EMA [45]. The intervention was intended for use as a stand-alone or adjunctive treatment or relapse prevention tool.

Initial usability testing of the intervention with 29 key stakeholders, including past or current gamblers, gambling clinicians, and gambling researchers, promisingly revealed high ratings of expected helpfulness for managing cravings and usability in quantitative and qualitative assessments [40]. Testing of the intervention is now needed under its conditions of intended use, that is, in real-time, real-world contexts by people seeking help for gambling problems.

Objectives

This study primarily aims to examine the acceptability and feasibility of GamblingLess: Curb Your Urge. The secondary aim of this study is to explore the intervention's preliminary effectiveness, with the intervention hypothesized to (1) reduce real-time gambling craving intensity from immediately before to immediately after using the intervention and (2) reduce the preintervention levels of gambling symptom severity, gambling cravings, gambling frequency and expenditure, craving self-efficacy, and gambling self-efficacy at postintervention and 1-month follow-up evaluations.

Methods

Trial Design

This study was a single-arm, 5-week acceptability and feasibility trial (1-week baseline and 4-week intervention periods) involving EMAs delivered 3 times daily and web-based evaluations at baseline, postintervention, and a 1-month follow-up.

Recruitment and Eligibility Criteria

The sample comprised 36 participants (14 women and 22 men) who completed the baseline measures, of whom 27 (82%) participants downloaded the app, 22 (61%) participants completed the postintervention evaluation, and 21 (58%) participants completed the follow-up evaluation. Individuals were eligible for inclusion if they were aged 18 years and above, resided in Australia, owned a smartphone, and were seeking help for their own gambling problem. Participants were permitted to access other help services throughout this study, with a mean number of 2.32 (SD 4.16) and 0.67 (SD 1.77) instances of additional professional help-seeking (eg, counseling, helpline, and general practitioners) reported by participants at the postintervention and follow-up evaluations, respectively. Participants were recruited across Australia via advertisements, using the tagline "Want to curb your gambling urges? And do it on your own? Then GamblingLess: Curb Your Urge may be for you," posted in gaming venues and on the web (eg, paid and free social media advertisements) and media announcements (print, radio, and television). In addition, counselors from several Australian gambling treatment services provided information about the study to clients seeking additional support.

The GamblingLess: Curb Your Urge Program

On the basis of the initial usability testing results [40], the research team refined the app intervention for this trial. Specifically, the intervention content was expanded to provide additional psychoeducation, mindfulness, and relaxation-based activities, and the tone was refined to be more normalizing. The intervention's esthetics and level of engagement were also improved through the use of color and additional audiovisual components, which were subtitled to ensure that they could be completed anywhere and anytime.

In this trial, GamblingLess: Curb Your Urge consisted of a 4-week EMI delivered by an existing smartphone app platform (MetricWire). The intervention comprised 12 urge-curbing tips and activities (eg, *About My Urge*, *Delay and Distract*, and *Urge Surfing*) described in Table S1, with examples presented in Figure S1 in Multimedia Appendix 1. Each tip and activity took 1 to 5 minutes to complete.

EMA Feature

The app administered a brief EMA at random times during three prespecified periods (9 AM to noon, 1 PM-4 PM, and 5 PM-8 PM) via push notifications. The EMA comprised five core items to measure gambling episodes (since the last EMA), gambling cravings (current and since the last EMA), gambling self-efficacy (current confidence in ability to limit or stop gambling), and craving self-efficacy (current confidence in the ability to resist cravings) and up to an additional seven items depending on core item responses. These additional items were included to gain descriptive data about gambling cravings (eg, duration, frequency) that are largely absent in the corresponding empirical literature [43]. The EMA took approximately 1-2 minutes to complete, depending on the pattern of responses (the EMA items are given in Table S2 in Multimedia Appendix 1).

EMI Feature

The app's EMI feature involved an automatic recommendation to use any urge-curbing tip or activity (Figure S2 in Multimedia Appendix 1) at the end of any EMA in which participants reported that they had a current craving (ie, responded *Yes* to EMA item 1; Table S2 in Multimedia Appendix 1). The program was also available 24×7 for use on demand. Whenever participants used a tip or activity (regardless of whether use was recommended or on demand), they were asked to rate the intensity of their craving to gamble from 0 (mild) to 10 (severe) immediately before and after using the tip or activity.

Procedure

Study advertisements directed participants to the web-based baseline questionnaire, which was preceded by the provision of study information (eg, what participation involves and a description of the app intervention) followed by participant consent. In the baseline questionnaire, participants provided their contact details, which were used to email participants an instruction manual for downloading and using the app (eg,

submitting EMAs and accessing the intervention content). Once enrolled, participants completed a 1-week EMA-only baseline period, followed by a 4-week intervention period (ie, EMA or EMI and 24×7 on-demand access to the intervention content). At the end of the intervention period, participants could no longer access the intervention. Participants were emailed a link to the web-based postintervention and 1-month follow-up questionnaires and, if required, were emailed up to 3 times to encourage questionnaire completion. Participants were recruited from September 2019 to June 2020. Participants were reimbursed up to Aus $60 (US $46.2) in e-gift vouchers, paid upon completion of the postintervention (Aus $30 [US $23.30]) and follow-up (Aus $30 [US $23.30]) evaluations. Reimbursement was not contingent on EMA completion. This study received ethics approval from the Deakin University Human Research Ethics Committee (Ethics ID: 2019-030).

Measures

Baseline, postintervention, and follow-up evaluation measures were administered via structured web-based questionnaires hosted by Qualtrics (10 min to complete). During-intervention measures were available via the MetricWire app. An overview of the assessment measures and time points are given in Table S3 in Multimedia Appendix 1.

Acceptability Measures

Acceptability was assessed via a number of measures administered at postintervention, including (1) helpfulness of each urge-curbing tip or activity, and the relevance and burden of EMA items, via separate 11-point Visual Analogue Scales (VAS) from 0 (not helpful or relevant or burdensome) to 10 (very helpful or relevant or burdensome) [40]; (2) satisfaction with the intervention via the 3-item Client-Satisfaction Questionnaire-3 (CSQ-3; total score range 3-12) [46], with higher scores indicating greater overall satisfaction, participant needs met, and likelihood of future participant use; (3) impact of the intervention on participants' awareness, knowledge, attitude, intention to change, help-seeking behavior, and behavior change in relation to gambling cravings via the 6-item App-Specific subscale of the Mobile App Rating Scale (MARS) [47], whereby the mean item score of 3 (range 1-5) indicates minimum acceptability [48]; and (4) a series of open-ended items assessing suggested improvements for any tip or activity that participants rated a 5 or less (out of 10) for helpfulness (*How could [tip/activity] be improved?*), any technical issues (*Please comment on any technical issues experienced.*), and general feedback about the app intervention (*Do you have any other comments or feedback?*) [40].

Feasibility Measures

Feasibility was assessed via participant recruitment at baseline and retention at postintervention and follow-up evaluations. Feasibility was also assessed using several app use metrics, including EMA compliance and EMI compliance. EMA compliance was measured as the rate at which participants completed the EMA during the baseline and intervention periods. EMI compliance was measured as the rate at which participants completed any intervention content within 60 minutes of receiving an EMI recommendation to use the intervention because they reported a current craving to gamble during an EMA [38,49,50]. Feasibility was also assessed via intervention use more generally. Upon inspection of the data, *any intervention use* was stratified by *EMA-prompted use* (defined as intervention use within 60 min of completing an EMA, regardless of whether they were recommended an activity based on craving occurrence) and *on-demand use* (defined as any other intervention use). The term *EMI use* was employed for the subset of EMA-prompted use in which use occurred following an EMI recommendation to use an intervention based on craving occurrence (note that this is equivalent to the EMI compliance rate).

Preliminary Effectiveness Measures

Preliminary effectiveness was assessed using changes in (1) real-time craving intensity measured by the aforementioned rating items administered immediately before and after using any urge-curbing tip or activity and (2) outcome measures completed at baseline, postintervention, and follow-up evaluations.

In the outcome evaluation, past-week gambling symptom severity (primary outcome) was measured using the 12-item Gambling-Symptom Assessment Scale (G-SAS) [51], with total scores (range 0-48) indicating mild (score of 8-20), moderate (score of 21-30), severe (score of 31-40), and extreme (score of 41-48) severity. In terms of secondary outcomes, gambling cravings were measured using the G-SAS Urge Subscale, which comprises the first 4 G-SAS items (score range 0-16) assessing the craving intensity, frequency, duration, and subjective control. Past-month gambling frequency (days) and expenditure (Aus $) on 6 gambling activities (electronic gaming machines or pokies, table games, racing, sports and event betting, number games, and informal private betting) were measured using a series of single items. Participants' confidence in their ability to resist a craving to gamble (*craving self-efficacy*) and to limit or stop their gambling (*gambling self-efficacy*) were measured using separate 11-point VAS from 0 (not at all) to 10 (very confident) [43].

Descriptive and Diagnostic Measures

At baseline, participants reported their demographic information and whether they had a problem with 6 gambling activities (eg, table games and number games). Participants also completed the 9-item Problem Gambling Severity Index (PGSI) [52] to measure their past-year problem gambling status. Total PGSI scores (range 0-27) were used to indicate nonproblem (score of 0), low-risk (score of 1-2), moderate-risk (score of 3-7), and problem (score of 8 or higher) gambling.

Statistical Analysis

Statistical analyses were conducted in Stata v13.0 [53]. Acceptability and feasibility were explored using descriptive statistics for quantitative variables (means and SDs for continuous variables and count and percentages for categorical data) and thematic content analysis at a semantic level for qualitative variables (ie, focusing on participant responses rather than latent meanings) [54]. Baseline differences in participants who did not complete a postintervention evaluation were

calculated using chi-square tests with Fisher exact *P* values for categorical variables and *t* tests for continuous variables.

For the preliminary effectiveness evaluation using smartphone data, there were no missing data, as responses to all app-based items were required. To explore the intervention's effectiveness for reducing real-time craving intensity, measured immediately before and immediately after intervention use via in-built rating items, a series of paired samples *t* tests using cluster-robust standard errors (to account for multiple intervention events clustered within individuals) assessed the change in mean intensity ratings. Separate *t* tests explored any intervention use and stratifications of EMA-prompted use (including EMI use) and on-demand use (defined in the Measures section). Given that some participants used multiple interventions in a short period because they could use the intervention as many times as they wanted, an additional *t* test assessed changes in craving intensity from immediately before the *first* intervention used to immediately after the *last* intervention used in any 60-minute window following an EMA. Owing to the pilot nature of this study and the small sample size, the unique effectiveness of each urge-curbing tip and activity was not examined.

For preliminary effectiveness outcome data collected at baseline, postintervention, and follow-up evaluations, there was no missingness across outcomes, except gambling expenditure (approximately 2.3% at baseline). Given the low amount of missing data and small sample size, analyses used a pairwise inclusion approach. To explore the intervention's preliminary cumulative effectiveness over the study period, a series of paired samples *t* tests assessed group-level changes in mean outcome scores from baseline to postintervention and baseline to follow-up. Cohen *d* effect sizes were interpreted as small (0.20), medium (0.50), and large (0.80) [55]; Cohen *d* was not calculated for total gambling frequency and expenditure because of their skewed nature. Group-level examinations of effectiveness were also supplemented by 2 metrics of individual-level change in G-SAS gambling symptom severity, G-SAS gambling cravings, gambling frequency, gambling expenditure, craving self-efficacy, and gambling self-efficacy from baseline to postintervention and baseline to follow-up. First, a Reliable Change Index (RCI) [56] assessed change beyond that attributable to measurement error or chance, where RCI1.96 indicates reliable change with 95% confidence [57]. Second, clinically significant change was calculated using functional score ranges where possible (G-SAS score of 20 or less) or a convention of at least a 25% change in scores in the positive direction [58]. Four categories of change were created: *recovered* (the final score indicated a reliable change and was in the functional range), *improved* (the final score indicated a reliable change but was in the dysfunctional range), *unchanged* (the final score did not indicate a reliable change), or *deteriorated* (the final score indicated a reliable change in the negative direction).

Results

Sample Descriptive Statistics

Sample descriptive statistics for the 36 participants who completed the baseline measures are presented in Table 1. The majority of participants were men (22/36, 61%), in the age range of 35 to 49 years (17/36, 47%), worked full time (25/36, 69%), and used an iOS operating system (20/36, 56%). Majority of the participants identified having a problem with informal private betting (n=34; 94%) and electronic gaming machines (27/36, 75%). Almost all of the participants (34/36, 94%) met the PGSI criteria for problem gambling, with the remainder (2/36, 6%) displaying moderate-risk gambling.

Table 1. Descriptive statistics for the overall sample at baseline (N=36).

Demographic and diagnostic measures	Value, n (%)
Gender (male)	22 (61)
Age group (years)	
18-24	1 (3)
25-34	14 (39)
35-49	17 (47)
50-64	4 (11)
65+	0 (0)
Born in Australia	11 (31)
Smartphone operating system	
Android	16 (44)
iOS (iPhone)	20 (56)
Employment	
Work full time	25 (69)
Work part time or casual	6 (17)
Unemployed	2 (6)
Full-time student	1 (3)
Full-time home duties	0 (0)
Retired	1 (3)
Other (work cover because of injury)	1 (3)
Past-month problem with gambling activities[a]	
Electronic gaming machines or pokies	27 (75)
Table games (eg, roulette and poker)	7 (19)
Horses, harness racing, or grayhound racing	15 (42)
Sports and event betting	14 (39)
Number games (eg, lotteries, keno, and bingo)	5 (14)
Informal private betting	34 (94)
Past-year problem gambling severity (Problem Gambling Severity Index)[b]	
No-risk gambling (score of 0)	0 (0)
Low-risk gambling (scores of 1-2)	0 (0)
Moderate-risk gambling (scores of 3-8)	2 (6)
Problem gambling (scores of 8 or higher)	34 (94)
Hazardous alcohol use (Alcohol Use Disorders Identified Test-3)[c]	28 (78)
High psychological distress (Distress thermometer)[d]	29 (81)

[a]Participants could indicate a problem with more than one gambling activity.

[b]The Problem Gambling Severity Index was used, with scores ranging from 0 to 27.

[c]The Alcohol Use Disorders Identified Test-3 [59] was used to measure hazardous alcohol use, defined as a score of 1 or more (range 0-4).

[d]The Distress thermometer was used to measure psychological distress, defined as a score of 4 or more (range 0-10) [60].

Acceptability

Descriptive acceptability statistics, presented in Table 2, were based on the 22 participants who completed the postintervention evaluation.

Table 2. Descriptive statistics for postintervention measures (n=22).

Acceptability measure	Value, mean (SD)
Helpfulness ratings for "urge-curbing tips and activities"[a]	
Tip—Delay and Distract	7.41 (2.06)
Tip—About My Urge	6.73 (2.39)
Activity—Change Your Thoughts	6.55 (2.46)
Activity—Get to Know Your Thoughts	6.55 (2.72)
Activity—Urge Surfing	6.55 (2.79)
Activity—Mindfulness	6.41 (2.52)
Activity—Breathing Relaxation	6.23 (2.88)
Tip—Tying it All Together	6.18 (2.56)
Tip—Talk to Someone	6.18 (2.72)
Activity—Progressive Muscle Relaxation	5.64 (2.66)
Activity—Belly Breathing	5.55 (2.74)
Activity—Brief Imagery	5.18 (2.89)
Ecological momentary assessment items[a]	
Relevance of items	7.45 (2.02)
Burdensome nature of items	4.59 (3.74)
Satisfaction with intervention (CSQ-3[b])	
Overall satisfaction with the intervention[c]	3.00 (0.69)
The intervention met my needs[c]	2.82 (1.01)
I would use the intervention again, if needed[c]	3.05 (0.79)
Total satisfaction with intervention[d]	8.86 (2.05)
Impact of the intervention (Mobile App Rating Scale app–specific subscale)[e]	
Awareness of the importance of addressing cravings	3.91 (1.15)
Help seeking in future for cravings	3.82 (1.33)
Knowledge and understanding of cravings	3.73 (1.16)
Attitudes toward addressing cravings	3.68 (0.99)
Intention to address cravings	3.64 (1.00)
Behavior change: the app would help to manage cravings	3.64 (1.09)

[a]Mean scores can range from 0 to 10; tips and activities are presented in descending order from highest to lowest mean helpfulness rating.

[b]CSQ-3: Client-Satisfaction Questionnaire-3.

[c]The mean CSQ-3 item scores ranged from 1 to 4.

[d]The mean CSQ-3 total scores ranged from 3 to 12.

[e]The mean Mobile App Rating Scale scores can range from 1 to 5; items are presented in descending order, from the highest to lowest mean rating.

Intervention Helpfulness and EMA Relevance and Burden

The mean ratings for the perceived helpfulness of all urge-curbing tips and activities were higher than 5 out of 10, indicating above-average helpfulness. The highest rated overall were Delay and Distract (mean 7.41, SD 2.06), About My Urge (mean 6.73, SD 2.39), Change Your Thoughts (mean 6.55, SD 2.46), Get to Know Your Thoughts (mean 6.55, SD 2.72), and Urge Surfing (mean 6.55, SD 2.79). The mean ratings of the EMA's perceived relevance were high (7.45 out of 10, SD 2.02) and burdensome nature were average (4.59 out of 10, SD 3.74).

Satisfaction With the Intervention

On the CSQ-3, participants scored an average of 8.86 out of 12 (SD 2.05). Individual item mean scores of approximately 3 out of 4 indicated that participants were mostly satisfied with the intervention, that the intervention met most of their needs, and that they would likely use it again to manage their cravings.

Impact of the Intervention

Mean scores across the MARS app–specific subscale items (range 3.64-3.91 out of 5) met acceptability standards (score of 3). From highest to lowest rated impact, participants indicated that the intervention improved their awareness about the importance of addressing cravings, future help seeking for cravings, knowledge and understanding of cravings, attitudes toward addressing cravings, and intention to address and ability to manage cravings.

Suggested Improvements

Overall, 16 participants rated the helpfulness of at least one tip or activity a 5 or below (out of 10) for managing cravings and were subsequently asked to suggest improvements on that tip or activity. Most of these participants tended to give a variation of "unsure" or no improvements required. Nevertheless, 1 participant suggested that Delay and Distract, Mindfulness, Urge Surfing, and Change Your Thoughts could be improved by providing more examples of "what has worked for others". Furthermore, 1 participant suggested improving Talk to Someone by adding a live chat function to enable in-the-moment messaging support and 2 participants suggested improving the mindfulness- and relaxation-based activities by linking them to "a proper meditation/relaxation program" and incorporating additional audiovisual components.

Technical Issues

The majority of participants (15/22, 68%) reported no technical issues with the intervention. Five participants reported that the audiovisual content was slow to load, 1 participant reported that the app exceeded their phone's storage capacity, and 1 participant reported that the tips and activities incorrectly appeared as *not completed* after completion. Almost all of the participants (20/22, 91%) reported no EMA technical issues; however, 1 participant reported that they did not receive notifications and 1 participant reported that EMAs wrongly displayed as *not completed* after completion.

General Feedback

Half of the participants (11/22, 50%) provided general feedback about the intervention, with most of them (7/22, 32%) providing positive feedback. Specifically, 3 participants found the EMA notifications helpful daily reminders to *stay on track and not gamble* and 4 participants found the intervention content to be highly accessible and helpful, with 1 participant stating that the app was "a great way to have different tools available at any time" and another stating that "the app worked well, was easy to use, notifications were good, (and the) training things were helpful and informative." Four participants provided negative feedback about the EMA notifications, whereby 2 participants requested tailoring capabilities to user-specified times (eg, "pay day"), 1 participant stated that the notifications "reminded me about gambling when I wasn't thinking about it," and 1 participant thought that the notifications were too frequent.

Feasibility

Consistent with the acceptability statistics, the feasibility statistics are based on the 22 participants who completed the postintervention evaluation, with the exception of the *Recruitment and Retention* section that details participant numbers throughout the study.

Recruitment and Retention

Over the 1-month recruitment period, 56 gamblers consented to participate in this study. Of these, 36 gamblers completed the baseline measures, 14 gamblers did not complete any baseline measures, and 6 gamblers completed only the baseline demographic measures. Of the 36 participants who completed the baseline measures, 9 participants did not download the app, 1 participant downloaded the app but did not use it, and 4 participants did not use the app beyond the first week of the study and did not complete the postintervention evaluation. The remaining 22 participants used the app and completed the postintervention evaluation, with 21 of these participants completing the 1-month follow-up evaluation. Of the 36 participants who completed baseline measures, 61% (22/36) completed postintervention evaluations and 58% (21/36) completed follow-up evaluations. There were no significant differences in any baseline variable between participants who did or did not complete the postintervention evaluation.

EMA Compliance

The EMAs had a compliance rate of 68% in the baseline period (mean 14.27, SD 5.68 out of 21 EMAs), 47% in the intervention period (mean 54.05, SD 33.37 out of 84 EMAs), and 51% overall (mean 54.05, SD 33.37 out of 105 EMAs).

EMI Compliance

The EMI compliance rate was 15%, as participants used an intervention 13 out of the 87 times that they were recommended to do so based on craving occurrence.

Intervention Use

Overall, 19 participants used the intervention at least once during the intervention period. Table 3 presents the descriptive statistics for intervention use. The intervention was used a total of 166 times (median 7; range 1-33), including 59 EMA-prompted uses (defined as intervention use within 60 min of completing an EMA), of which 13 were EMI uses (defined as intervention use within 60 min of completing an EMA and following an EMI recommendation to use an intervention because the participant reported a current craving). Of the 59 EMA-prompted uses, participants used the intervention once between EMAs on 29 occasions, 2 times on 10 occasions, 3 times on 2 occasions, and 4 times on 1 occasion. In contrast, there were 107 on-demand uses (defined as any other intervention use). The most used intervention content included About My Urge (22 uses), Talk to Someone (19 uses), and Mindfulness (18 uses).

Table 3. Total intervention use frequencies (count and percentages) stratified by ecological momentary assessment–prompted and on-demand use during the 4-week intervention period (n=22).

Intervention content	Total use, n (%)	Ecological momentary assessment–prompted use[a], n (%)	On-demand use[b], n (%)
Tip—About My Urge	22 (13.3)	7 (11.9)	15 (14.0)
Tip—Talk to Someone	19 (11.4)	9 (15.3)	10 (9.3)
Activity—Mindfulness	18 (10.8)	7 (11.9)	11 (10.3)
Activity—Get to Know Your Thoughts	16 (9.6)	4 (6.8)	12 (11.2)
Activity—Change Your Thoughts	13 (7.8)	1 (1.7)	12 (11.2)
Tip—Tying it All Together	13 (7.8)	6 (10.2)	7 (6.5)
Activity—Belly Breathing	12 (7.2)	8 (13.6)	4 (3.7)
Activity—Urge Surfing	12 (7.2)	2 (3.4)	10 (9.3)
Activity—Brief Imagery	11 (6.6)	4 (6.8)	7 (6.5)
Tip—Delay and Distract	11 (6.6)	5 (8.5)	6 (5.6)
Activity—Breathing Relaxation	10 (6.0)	3 (5.1)	7 (6.5)
Activity—Progressive Muscle Relaxation	9 (5.4)	3 (5.1)	6 (5.6)
Total	166 (100.0)	59 (100.0)	107 (100.0)

[a]Ecological momentary assessment–prompted use is defined as intervention use within 60 min of completing an ecological momentary assessment, regardless of whether participants were recommended an activity based on craving occurrence.

[b]On-demand use is defined as any other intervention use.

Preliminary Effectiveness

Descriptive Statistics

On the basis of the 22 participants who completed the postintervention evaluation, Table 4 presents descriptive statistics for EMA outcome variables. Compared with the 1-week baseline period, there was a 71% reduction in the average number of gambling episodes and a 72% reduction in the average number of craving occurrences, as reported in the 4-week intervention period. Furthermore, craving self-efficacy and gambling self-efficacy increased from mean ratings of 5.85 (SD 2.95) and 4.82 (SD 3.01) out of 10 (where 10=complete confidence in ability to resist a craving and to limit or stop gambling) over the baseline period to mean ratings of 6.63 (SD 2.61) and 5.72 (SD 3.02) over the intervention period, respectively.

Table 4. Descriptive statistics of ecological momentary assessment outcome variables (n=22).

Variable	One-week baseline period	Four-week intervention period	Total
Gambling episodes, n (%)			
No	304 (86.1)	835 (93.6)	1139 (91.5)
Yes	49 (13.9)	57 (6.4)	106 (8.5)
Gambling episode win or loss status, n (%)			
Win	10 (20.4)	8 (14.0)	18 (17.0)
Loss	33 (67.3)	42 (73.7)	75 (70.7)
Broke even	6 (12.2)	7 (12.3)	13 (12.3)
Gambling episode win amount, Aus $ (US $), n (%)			
1-150 (0.70-115.6)	6 (60.0)	2 (25.0)	8 (44.4)
151-500 (116.4-385.3)	2 (20.0)	2 (25.0)	4 (22.2)
501-1000 (386.1-770.6)	1 (10.0)	3 (37.5)	4 (22.2)
1001-7500 (771.3-5779.2)	1 (10.0)	1 (12.5)	2 (11.1)
Gambling episode loss amount, Aus $ (US $), n (%)			
1-150 (0.70-115.6)	9 (27.3)	12 (28.6)	21 (28.0)
151-500 (116.4-385.3)	15 (45.4)	7 (16.7)	22 (29.3)
501-1000 (386.1-770.6)	5 (15.2)	15 (35.7)	20 (26.7)
1001-7500 (771.3-5779.2)	4 (12.1)	8 (19.0)	12 (16.0)
Craving occurrences, n (%)			
No	274 (77.6)	805 (90.2)	1079 (86.7)
Yes	79 (22.4)	87 (9.8)	166 (13.3)
Craving intensity[a], mean (SD)	6.35 (3)	6.59 (3)	6.47 (3)
Craving frequency[b], mean (SD)	4.22 (9)	5.39 (16)	4.78 (13)
Craving duration (min)[c], mean (SD)	36.64 (59)	50.65 (79)	43.36 (70)
Subjective control over cravings[a], mean (SD)	5.96 (4)	5.95 (4)	5.96 (4)
Craving self-efficacy[a], mean (SD)	5.85 (3)	6.63 (3)	6.41 (3)
Gambling self-efficacy[a], mean (SD)	4.82 (3)	5.71 (3)	5.46 (3)

[a]Range 0-10 on a Visual Analogue Scale.

[b]Range 1-180 craving occurrences.

[c]Range 0-480 minutes.

Real-Time Reduction in Craving Intensity Immediately After Intervention Use

On the basis of the 22 participants who completed the postintervention evaluation, the results of clustered paired-sample *t* tests revealed a significant decrease of 5.4% in momentary craving intensity from immediately before to immediately after *any intervention use* (P=.01), with a medium effect (Cohen d_z=−0.64; 95% CI −1.13 to −0.14). The results demonstrated a 7.5% decrease in craving intensity for EMA-prompted use (P=.03; Cohen d_z=−0.72; 95% CI −1.34 to −0.06), and more specifically, a 10.5% decrease for EMI use (P=.01; Cohen d_z=−1.29; 95% CI −2.48 to −0.03). In contrast, there was a 4.5% decrease for on-demand use (P=.01; Cohen d_z=−0.66; 95% CI −1.16 to −0.14). There was no added benefit of using multiple interventions between EMAs, as the results demonstrated an 8.6% decrease in craving intensity after one use (P=.02; Cohen d_z=−0.94; 95% CI −1.67 to −0.17) compared with a 7.7% decrease after more than one use (P=.29; Cohen d_z=−0.63; 95% CI −1.49 to 0.28).

Change in Gambling Symptom Severity, Cravings, Frequency, Expenditure, and Self-Efficacy at Postintervention and Follow-Up

Descriptive statistics for gambling symptom severity, cravings, frequency, expenditure, and self-efficacy measured at baseline (n=36), postintervention (n=22), and follow-up (n=21) are presented in Table 5. At the group level, participants showed significant reductions in mean G-SAS gambling symptom severity (d=0.61 and 0.75; P=.01 and .003), G-SAS cravings (d=0.49 and 0.55; P=.03 and .02), gambling frequency (P=.01

and .004), and gambling expenditure (P=.04 and .003) from baseline to postintervention and follow-up, respectively. In addition, participants showed a significant increase in mean craving self-efficacy and gambling self-efficacy from baseline to postintervention (d=0.66 and 0.60; P=.01 and .01) and gambling self-efficacy (d=0.49; P=.04) but not craving self-efficacy (P=.20) from baseline to follow-up.

Table 5. Descriptive statistics (mean and SD) for measures administered at baseline, postintervention, and 1-month follow-up evaluations.

Outcome measure	Baseline (N=36)	Postintervention (n=22)	One-month follow-up (n=21)
Past-week gambling symptom severity[a] (Gambling-Symptom Assessment Scale), mean (SD)	30.92 (8)	22.18 (10)	20.43 (11)
Mild, n (%)	5 (14)	9 (41)	11 (52)
Moderate, n (%)	11 (31)	7 (32)	6 (29)
Severe, n (%)	14 (39)	6 (27)	4 (19)
Extreme, n (%)	6 (17)	0 (0)	0 (0)
Past-week gambling cravings[b] (Gambling-Symptom Assessment Scale-Urge Subscale), mean (SD)	9.81 (3)	6.82 (3)	6.57 (4)
Total gambling frequency (days) in the last month, mean (SD)	14.22 (13)	5.27 (6)	3.50 (5)
Total gambling expenditure (Aus $) in the last month, mean (SD)	2894.17 (3736)	1675.68 (2484)	629.48 (1232)
Current gambling self-efficacy[c], mean (SD)	3.97 (3)	6.32 (3)	6.00 (3)
Current craving self-efficacy[c], mean (SD)	4.72 (3)	6.77 (2)	6.00 (3)
Gambling treatment goal, n (%)			
Quit gambling altogether	24 (67)	14 (64)	14 (67)
Quit gambling activities I think I have a problem with	8 (22)	5 (23)	5 (24)
Cut back gambling activities I think I have a problem with	4 (11)	3 (14)	2 (10)
Additional professional help-seeking for gambling problems (number of times) in the last month, mean (SD)	N/A[d]	2.32 (4)	0.67 (2)

[a]Gambling-Symptom Assessment Scale scores can range from 0 to 48, categorized as mild (score of 8-20), moderate (score of 21-30), severe (score of 31-40), and extreme (score of 41-48).

[b]Gambling-Symptom Assessment Scale-Urge Subscale scores can range from 0 to 16.

[c]Self-efficacy scores can range from 0 to 10 on a Visual Analogue Scale.

[d]N/A: not applicable.

Table 6 presents clinically significant change on these measures from baseline to postintervention and follow-up. At an individual level, approximately a quarter of participants were *recovered* at postintervention and follow-up, respectively, on G-SAS gambling symptom severity (6/22, 27%; 6/21, 29%) and G-SAS gambling cravings (6/22, 27%; 7/21, 33%), the majority were *unchanged* (symptom severity: 13/22, 59%; 11/21, 52%; cravings: 14/22, 64%; 14/21, 67%), and a small number of participants were *deteriorated* at postintervention (symptom severity: 1/22, 5%; cravings: 2/21, 9%); however, none were deteriorated at follow-up. The majority of participants were *unchanged* at postintervention and follow-up on gambling frequency (20/22, 91%; 18/21, 90%), gambling expenditure (19/22, 86%; 17/21, 81%), craving self-efficacy (18/22, 82%; 17/21, 81%), and gambling self-efficacy (18/22, 82%; 18/21, 86%); however, none were *deteriorated* on these measures.

Table 6. Clinically significant changes in outcome measures at postintervention (n=22) and 1-month follow-up (n=21) evaluations.

Outcome measure	Postintervention (n=22)	One-month follow-up (n=21)
Gambling symptom severity (G-SAS[a]), n (%)		
Recovered[b]	6 (27)	6 (28)
Improved	2 (9)	4 (19)
Unchanged	13 (59)	11 (52)
Deteriorated	1 (4)	0 (0)
Gambling cravings (G-SAS Urge Subscale), n (%)		
Recovered	6 (27)	7 (33)
Improved	0 (0)	0 (0)
Unchanged	14 (63)	14 (66)
Deteriorated	2 (9)	0 (0)
Total gambling frequency[c] (days), n (%)		
Recovered	2 (9)	2 (10)
Improved	0 (0)	0 (0)
Unchanged	20 (90)	18 (90)
Deteriorated	0 (0)	0 (0)
Total gambling expenditure (Aus $), n (%)		
Recovered	3 (13)	4 (19)
Improved	0 (0)	0 (0)
Unchanged	19 (86)	17 (80)
Deteriorated	0 (0)	0 (0)
Craving self-efficacy (11-point VAS[d]), n (%)		
Recovered	0 (0)	0 (0)
Improved	4 (18)	3 (14)
Unchanged	18 (81)	17 (80)
Deteriorated	0 (0)	1 (4)
Gambling self-efficacy (11-point VAS), n (%)		
Recovered	0 (0)	0 (0)
Improved	4 (18)	3 (14)
Unchanged	18 (81)	18 (85)
Deteriorated	0 (0)	0 (0)

[a]G-SAS: Gambling-Symptom Assessment Scale.

[b]Recovered: the final score indicated a reliable change and was in the functional range, indicated by a score of 20 or less on the Gambling-Symptom Assessment Scale or at least a 25% reduction in scores for gambling cravings, total gambling frequency, and total gambling expenditure, and at least a 25% increase in scores for craving self-efficacy and gambling self-efficacy, in postintervention and follow-up evaluations.

[c]n=20 at 1-month follow-up, as 1 participant did not report gambling frequency.

[d]VAS: Visual Analogue Scale.

Discussion

Principal Findings

The results of this study demonstrate the acceptability, feasibility, and preliminary effectiveness of the first smartphone app–delivered EMI (GamblingLess: Curb Your Urge), which aims to prevent gambling episodes through reduced craving intensity in people seeking help for gambling problems.

Acceptability

Overall, the app intervention was rated favorably across quantitative and qualitative measures of acceptability. Participants indicated that they were mostly satisfied with the intervention, as it met most of their needs and could be used to

manage their cravings in the future, based on mean CSQ-3 scores. Furthermore, participants indicated that the intervention improved their awareness, knowledge, attitude, and intentions to address cravings and, importantly, their ability to manage cravings, based on mean MARS item scores. Qualitatively, the majority of participants provided positive feedback about the intervention, particularly in relation to its accessibility, helpfulness, and effectiveness in preventing gambling episodes.

Promisingly, participants rated all of the urge-curbing tips and activities as above average for helpfulness in managing cravings. Consistent with the evidence base for gambling treatments, in which CBT techniques are considered best practice [7,41,61,62], the highest rated content involved CBT techniques, including psychoeducation (About My Urge); cognitive reappraisal (Get to Know Your Thoughts and Change Your Thoughts); and relapse prevention techniques, including distraction (Delay and Distract) and a mindfulness practice to *surf* the wave of a craving without acting on it (Urge Surfing). Participants suggested that this content could be further improved by providing more concrete examples of skill implementation. The lowest rated content included relaxation-based or imaginal techniques (Progressive Muscle Relaxation, Belly Breathing, and Brief Imagery), with 2 participants suggesting their improvement via linkage with related programs outside of the app. Taken together, these findings suggest that intervention content variability and examples of skill implementation are important for people to find, and learn to apply, what works best for them.

With respect to the EMA, participants considered them highly relevant but somewhat burdensome, which represents an improvement from the initial usability testing [40] in which they were rated in the average range on both domains. To reduce its level of burden, the frequency of EMAs could be reduced or, as 2 participants suggested, made personally customizable to specific times (eg, *pay day*). Indeed, the combined results of usability testing and this trial indicated that tailoring capabilities were considered particularly important for consumers to enable personalized check-ins and treatment.

Feasibility

Recruitment and retention have posed a challenge in studies examining app-delivered interventions for mental health [61], particularly for problem gambling [29]. Despite extensive recruitment efforts, only 36 people completed baseline measures, with approximately 60% of these people completing postintervention (n=22) and follow-up (n=21) evaluations. Retention rates in this trial were lower than those of other feasibility trials of app-delivered EMIs for alcohol and substance use, where rates range from 63% to 90% [35,38]. Although not directly comparable, retention was on par with the only other feasibility trial of a gambling intervention app (60%) [26]. Future research may benefit from examining consumer profiles to enable targeted recruitment campaigns and tailored interventions to particular demographic groups, the latter of which may also facilitate trial retention through improved engagement with the app [62].

In this trial, the feasibility was measured via compliance with the EMA and EMI features. The overall compliance was 51% for EMAs and 15% for EMIs. Although there are no directly comparable gambling app intervention studies, previous gambling EMA studies have reported variable compliance rates from 50% to over 90% [43,63]; however, notably, EMA completion was financially incentivized in these studies. As financial incentives are not feasible for ongoing app interventions in real-world contexts, it would be useful to examine ways to improve EMA compliance. In addition, EMA compliance may have been reduced in this study as participants could access intervention content *on demand* without needing to complete a more time-consuming EMA. While this is a limitation of providing a hybrid *push or pull* intervention, this approach supports consumers to access interventions regardless of their awareness of vulnerable states or motivation to seek support [45]. Furthermore, EMA compliance may improve outside of a research context, as it is envisioned that the EMA feature of future iterations of the intervention for real-world use would be briefer (eg, to only measure current craving occurrence) and customizable regarding the frequency and timing of administration (eg, to user-specified times). The EMI compliance rate in this study was also low, however, as participants used the intervention only 13 of the 87 times that they were recommended to do so. Low EMI compliance may reflect the limitations of the app platform, whereby intervention content cannot be automatically delivered (rather than just recommended). Although this limitation enabled a more nuanced evaluation of the intervention's preliminary effectiveness under different conditions of use, future app iterations may benefit from using platforms that can automatically deliver an EMI at times of need.

Across the 4-week intervention period, participants used the intervention a total of 166 times (median 7 uses; range 1-33), including 59 EMA-prompted uses and 107 on-demand uses. This rate of intervention uptake (7 uses over 4 weeks) appears comparable with that of traditional face-to-face gambling services, in which treatment typically involves 1 session per week for 10 weeks [64]. Despite similar uptake rates, we might expect higher app use given its 24×7 availability and evidence of EMI underutilization. Although apps can overcome a number of barriers to accessing face-to-face services (eg, geographic constraints) [11,12], future research should explore app-specific barriers to use at times of identified need. The higher rate of on demand (*pull*) intervention use compared with EMA-prompted and EMI (*push*) use suggests either that the EMA item measuring craving occurrence was not adequately identifying the times of vulnerability and/or that participants knew when they needed to access an intervention and were motivated to do so [45]. To address the latter possibility, future app interventions may benefit from incorporating additional user-initiated EMAs [18,34], whereby participants can record a gambling craving or episode at any time in the app.

Preliminary Effectiveness

A unique strength of EMA and EMI data is the capability to explore real-time intervention effects in people's everyday lives. Promisingly, descriptive EMA data showed that, compared with the baseline period, there was an approximately 71% to 72% reduction in the average number of gambling episodes and craving occurrences and an approximately 1-point increase in the mean craving self-efficacy and gambling self-efficacy on

an 11-point VAS, reported in the intervention period. In addition, the results showed that the intervention led to a medium decrease of 5.4% in real-time craving intensity from immediately before to immediately after use. This effect increased with a large decrease of 10.5% for EMI use (ie, where participants had a current craving to gamble) and 7.5% for EMA-prompted use (ie, use within 60 min of an EMA regardless of whether participants had a craving) and reduced with a medium decrease of 4.5% for on-demand use (ie, any other intervention use), which intuitively supports the intervention's increased effectiveness in moments of vulnerability. The results also suggested an optimal dosage of 1 intervention between EMAs, as there was no added benefit of multiple uses.

Given the small sample size and limited power to detect significant effects, this study also explored the intervention's preliminary effectiveness at both the group and individual level. At the group level, participants displayed significant medium-to-large reductions in mean gambling symptom severity and cravings and reductions in gambling frequency and expenditure at postintervention and 1-month follow-up evaluations. Participants also displayed significant medium-to-large increases in gambling self-efficacy at postintervention and follow-up and craving self-efficacy at postintervention. At an individual level, more than a quarter of participants were considered *recovered* or *improved* on G-SAS gambling symptom severity (8/22, 36%; 10/21, 48%) and G-SAS gambling cravings (6/22, 27%; 7/21, 33%), at postintervention and follow-up respectively, with the majority remaining *unchanged*, which compares with more than a half of participants (n=55, 64%) on gambling symptom severity in a randomized trial of its *parent* program (GamblingLess) [41]. A small proportion of participants in this study were also *recovered* or *improved* on gambling frequency, expenditure, gambling self-efficacy, and craving self-efficacy (2/22, 9% to 4/22, 19%). By the end of this study, only 1 participant was considered *deteriorated* on craving self-efficacy but no other measure. Although it is possible that these results reflect a highly motivated sample, natural recovery, or recovery resulting from concurrent treatment, these findings provide preliminary evidence for the effectiveness of the intervention, which would likely increase further when combined with a fuller suite of gambling interventions.

Strengths and Limitations

This study has several important strengths. First, the intervention was based on sound theory [42] and empirical research [43] to provide an evidence-based, app-delivered gambling intervention that supports self-management in light of low professional treatment uptake rates [9]. Second, this study and a preceding usability study [40] demonstrated the value of including consumers' feedback in the early stages of intervention development and evaluation, which is integral to short- and long-term app engagement [21,65]. Finally, the use of smartphones in this study to deliver real-time interventions to people in their natural environments and the lack of financial incentives for EMA or EMI completion support real-world uptake of the app.

The study findings should be interpreted in light of several limitations. Although this study's sample size was consistent with other acceptability and feasibility trials [18,26,38], it was relatively small, which limits the generalizability of the results and power to detect significant effects. Future studies would benefit from using a larger sample, which may require a longer recruitment period and substantial recruitment efforts [29]. Somewhat low rates of EMA and EMI compliance and intervention use also limited our statistical power to detect significant effects. Consistent with participant acceptability feedback, these rates could be increased by improving the intervention's level of engagement, such as through increased content variability and customizable features [62]. In addition, preliminary evidence of the effectiveness of the intervention may have been inflated by assessment reactivity and demand characteristics, as participants could likely ascertain the research aims. Interestingly, a recent gambling EMA study found a weak reactivity effect in the opposite direction than expected, as participants reported more gambling activity throughout the study [44], which is consistent with the broader addiction literature, which did not indicate strong reactivity effects [66]. Nevertheless, future studies may improve the generalizability of their results by examining intervention effectiveness using publicly available, routinely collected data from app use in the real world [67]. Finally, it is acknowledged that this study did not redress the dearth of economic data to support the use of smartphone-delivered interventions in mental health treatment, which future feasibility studies ought to consider [68].

Conclusions

The findings of this study support the acceptability, feasibility, and preliminary effectiveness of GamblingLess: Curb Your Urge, the first app-delivered EMI for craving management in people seeking help for gambling problems. These findings indicate the utility of developing targeted, real-time app interventions for problem gambling, particularly as an offshoot of more comprehensive programs with a developing or established evidence base, such as GamblingLess [41]. In so doing, app interventions may extend the reach of evidence-based treatment beyond the confines of face-to-face services to moments of need in people's everyday lives.

Acknowledgments

Over the past 3 years, SM, GY, and ND have received funding from multiple sources, including government departments and the Victorian Responsible Gambling Foundation (through hypothecated taxes from gambling revenue) and the National Association for Gambling Studies (NAGS), a not-for-profit organization with individual members across all stakeholder groups, which derives its funding from member fees and conference proceedings. SM has been the Victorian state representative (unpaid) on the NAGS Executive Committee, which includes representatives from all stakeholder groups. ND is a recipient of the Deakin University

Faculty of Health Mid-Career Fellowship. None of the authors received research funding from the gambling industry or any industry-sponsored organization.

Authors' Contributions

CH was involved in conceptualization, data collection, formal analysis, project administration, and writing the original draft. SM, GY, and ND were engaged in conceptualization, formal analysis, and writing—review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
GamblingLess: Curb Your Urge: Intervention content, ecological momentary assessment items, an overview of evaluation measures, and the ecological momentary intervention feature. EMA: ecological momentary assessment; EMI: ecological momentary intervention.
[DOCX File , 1413 KB - jmir_v23i3e25786_app1.docx]

References

1. Neal P, Delfabbro P, O'Neil M. Problem gambling and harm: Toward a national definition. Melbourne: Office of Gaming and Racing; 2005. URL: https://www.gamblingresearch.org.au/sites/default/files/2019-10/Problem%20Gambling%20and%20Harm%20-%20Towards%20a%20National%20Definition%202005.pdf [accessed 2021-03-10]
2. Langham E, Thorne H, Browne M, Donaldson P, Rose J, Rockloff M. Understanding gambling related harm: a proposed definition, conceptual framework, and taxonomy of harms. BMC Public Health 2016 Jan 27;16(1):1-23. [doi: 10.1186/s12889-016-2747-0]
3. Dowling NA, Youssef GJ, Jackson AC, Pennay DW, Francis KL, Pennay A, et al. National estimates of Australian gambling prevalence: f indings from a dual-frame omnibus survey. Addiction 2015 Nov 06;111(3):420-435. [doi: 10.1111/add.13176]
4. Gainsbury SM, Russell A, Hing N, Wood R, Lubman DI, Blaszczynski A. The prevalence and determinants of problem gambling in Australia: Assessing the impact of interactive gambling and new technologies. Psychology of Addictive Behaviors 2014 Sep;28(3):769-779. [doi: 10.1037/a0036207]
5. Browne M, Greer N, Rawat V, Rockloff M. A population-level metric for gambling-related harm. International Gambling Studies 2017 Mar 31;17(2):163-175. [doi: 10.1080/14459795.2017.1304973]
6. Cowlishaw S, Merkouris S, Dowling N, Anderson C, Jackson A, Thomas S. Psychological therapies for pathological and problem gambling. Cochrane Database Syst Rev 2012 Nov 14;11:1-88. [doi: 10.1002/14651858.CD008937.pub2] [Medline: 23152266]
7. Thomas SA, Merkouris SS, Radermacher HL, Dowling NA, Misso ML, Anderson CJ, et al. Australian guideline for treatment of problem gambling: an abridged outline. Medical Journal of Australia 2011 Dec 12;195(11-12):664-665. [doi: 10.5694/mja11.11088]
8. Yakovenko I, Quigley L, Hemmelgarn BR, Hodgins DC, Ronksley P. The efficacy of motivational interviewing for disordered gambling: Systematic review and meta-analysis. Addictive Behaviors 2015 Apr;43:72-82. [doi: 10.1016/j.addbeh.2014.12.011]
9. Productivity C. Gambling, Report no. 50. Canberra. URL: https://www.pc.gov.au/inquiries/completed/gambling-2010/report [accessed 2021-03-10]
10. Suurvali H, Hodgins D, Toneatto T, Cunningham J. Treatment Seeking Among Ontario Problem Gamblers: Results of a Population Survey. Psychiatric Services 2008 Nov;59(11):1343-1346. [doi: 10.1176/ps.2008.59.11.1343]
11. Suurvali H, Cordingley J, Hodgins DC, Cunningham J. Barriers to Seeking Help for Gambling Problems: A Review of the Empirical Literature. Journal of Gambling Studies 2009 Jun 24;25(3):407-424. [doi: 10.1007/s10899-009-9129-9]
12. Gainsbury S, Hing N, Suhonen N. Professional Help-Seeking for Gambling Problems: Awareness, Barriers and Motivators for Treatment. J Gambl Stud 2013 Mar 15;30(2):503-519. [doi: 10.1007/s10899-013-9373-x]
13. Goslar M, Leibetseder M, Muench HM, Hofmann SG, Laireiter A. Efficacy of face-to-face versus self-guided treatments for disordered gambling: A meta-analysis. Journal of Behavioral Addictions 2017 Jun;6(2):142-162. [doi: 10.1556/2006.6.2017.034]
14. Carlbring P, Smit F. Randomized trial of internet-delivered self-help with telephone support for pathological gamblers. Journal of Consulting and Clinical Psychology 2008 Dec;76(6):1090-1094. [doi: 10.1037/a0013603]
15. Casey LM, Oei TPS, Raylu N, Horrigan K, Day J, Ireland M, et al. Internet-Based Delivery of Cognitive Behaviour Therapy Compared to Monitoring, Feedback and Support for Problem Gambling: A Randomised Controlled Trial. Journal of Gambling Studies) 2017 Jan 25;33(3):993-1010. [doi: 10.1007/s10899-016-9666-y]
16. Harrison V, Proudfoot J, Wee PP, Parker G, Pavlovic DH, Manicavasagar V. Mobile mental health: Review of the emerging field and proof of concept study. Journal of Mental Health 2011 Oct 11;20(6):509-524. [doi: 10.3109/09638237.2011.608746]

17. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23. [doi: 10.2196/mhealth.8873]
18. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, Acceptability, and Preliminary Efficacy of a Smartphone Intervention for Schizophrenia. Schizophrenia Bulletin 2014 Mar 08;40(6):1244-1253. [doi: 10.1093/schbul/sbu033]
19. Bouchard S, Robillard G, Giroux I, Jacques C, Loranger C, St-Pierre M, et al. Using Virtual Reality in the Treatment of Gambling Disorder: The Development of a New Tool for Cognitive Behavior Therapy. Frontiers in Psychiatry 2017 Feb 24;8:1-10. [doi: 10.3389/fpsyt.2017.00027]
20. Ledgerwood DM, Petry NM. What do we know about relapse in pathological gambling? Clinical Psychology Review 2006 Mar;26(2):216-228. [doi: 10.1016/j.cpr.2005.11.008]
21. Badawy SM, Shah R, Beg U, Heneghan MB. Habit Strength, Medication Adherence, and Habit-Based Mobile Health Interventions Across Chronic Medical Conditions: Systematic Review. Journal of medical Internet research 2020 Apr 28;22(4):e17883. [doi: 10.2196/17883]
22. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: A systematic review. Journal of telemedicine and telecare 2016 Oct 18;24(1):22-30. [doi: 10.1177/1357633x16673538]
23. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral Functionality of Mobile Apps in Health Interventions: A Systematic Review of the Literature. JMIR mHealth uHealth 2015 Feb 26;3(1):e20. [doi: 10.2196/mhealth.3335]
24. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for Smarter Delivery of Mental Health Programs: A Systematic Review. Journal of Medical Internet Research 2013 Nov 15;15(11):e247. [doi: 10.2196/jmir.2791]
25. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health—a systematic review and meta-analysis. NPJ Digital Medicine 2019 Dec 02;2(1):1-10. [doi: 10.1038/s41746-019-0188-8]
26. Pfund RA, Whelan JP, Meyers AW, Peter SC, Ward KD, Horn TL. The Use of a Smartphone Application to Complete Therapeutic Homework in Cognitive-Behavioral Therapy for Gambling Disorder: a Pilot Study of Acceptability and Feasibility. J. technol. behav. sci 2019 Nov 28;5(2):156-163. [doi: 10.1007/s41347-019-00123-9]
27. So R, Furukawa TA, Matsushita S, Baba T, Matsuzaki T, Furuno S, et al. Unguided Chatbot-Delivered Cognitive Behavioural Intervention for Problem Gamblers Through Messaging App: A Randomised Controlled Trial. ournal of Gambling Studies 2020 Mar 11;36(4):1391-1407. [doi: 10.1007/s10899-020-09935-4]
28. Humphrey G, Chu J, Dowling N, Rodda S, Merkouris S, Parag V, et al. Manaaki – a cognitive behavioral therapy mobile health app to support people experiencing gambling problems: a randomized control trial protocol. BMC Public Health 2020 Feb 06;20(1):1-11. [doi: 10.1186/s12889-020-8304-x]
29. Humphrey G, Newcombe D, Whittaker R, Parag V, Bullen C. SPGeTTI: a Smartphone-Based Problem Gambling Evaluation And Technology Testing Initiative Final Report National Institute for Health Innovation. Auckland, New Zealand: Auckland UniServices Limited. The University of Auckland; 2019. URL: https://www.health.govt.nz/system/files/documents/publications/20190424-spgetti-354913-00-final-report.pdf [accessed 2021-03-10]
30. Coral R, Esposito F, Weinstock J. Don't Go There: A Zero-Permission Geofencing App to Alleviate Gambling Disorders. 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC) 2020:1-6. [doi: 10.1109/ccnc46108.2020.9045251]
31. Heron K, Smyth J. Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. British Journal of Health Psychology 2010;15(1):1-39. [doi: 10.1348/135910709x466063]
32. Schueller SM, Aguilera A, Mohr DC. Ecological momentary interventions for depression and anxiety. Depression and anxiety 2017 May 11;34(6):540-545. [doi: 10.1002/da.22649]
33. Bell IH, Lim MH, Rossell SL, Thomas N. Ecological Momentary Assessment and Intervention in the Treatment of Psychotic Disorders: A Systematic Review. Psychiatric Services 2017 Nov;68(11):1172-1181. [doi: 10.1176/appi.ps.201600523]
34. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A Smartphone Application to Support Recovery From Alcoholism. JAMA Psychiatry 2014 May 01;71(5):566. [doi: 10.1001/jamapsychiatry.2013.4642]
35. Shrier LA, Rhoads A, Burke P, Walls C, Blood EA. Real-time, contextual intervention using mobile technology to reduce marijuana use among youth: A pilot study. Addictive Behaviors 2014 Jan;39(1):173-180. [doi: 10.1016/j.addbeh.2013.09.028]
36. Schulte M, Liang D, Wu F, Lan Y, Tsay W, Du J, et al. A Smartphone Application Supporting Recovery from Heroin Addiction: Perspectives of Patients and Providers in China, Taiwan, and the USA. Journal of Neuroimmune Pharmacology 2016 Feb 4;11(3):511-522. [doi: 10.1007/s11481-016-9653-1]
37. Barrio P, Ortega L, López H, Gual A. Self-management and Shared Decision-Making in Alcohol Dependence via a Mobile App: a Pilot Study. International Journal of Behavioral Medicine 2017 Feb 24;24(5):722-727. [doi: 10.1007/s12529-017-9643-6]
38. Dennis ML, Scott CK, Funk RR, Nicholson L. A Pilot Study to Examine the Feasibility and Potential Effectiveness of Using Smartphones to Provide Recovery Support for Adolescents. Substance Abuse 2015 Oct 15;36(4):486-492. [doi: 10.1080/08897077.2014.970323]

39. Khazaal Y, Monney G, Richter F, Achab S. « Jeu-contrôle », rationnel d'une application de soutien aux limites de jeux. Journal de Thérapie Comportementale et Cognitive 2017 Sep;27(3):129-137. [doi: 10.1016/j.jtcc.2017.05.003]
40. Merkouris SS, Hawker CO, Rodda SN, Youssef GJ, Dowling NA. GamblingLess: Curb Your Urge: Development and usability testing of a smartphone-delivered ecological momentary intervention for problem gambling. International Gambling Studies 2020 Apr 10;20(3):515-538. [doi: 10.1080/14459795.2020.1749293]
41. Dowling N, Merkouris S, Rodda S, Smith D, Lavis T, Lubman D, et al. Development and evaluation of an online gambling self-directed programffective integration into existing services. Melbourne: Victorian Responsible Gambling Foundation; 2018. URL: https://responsiblegambling.vic.gov.au/resources/publications/development-and-evaluation-of-an-online-gambling-self-directed-program-effective-integration-into-existing-services-348/ [accessed 2021-03-10]
42. Witkiewitz K, Marlatt G. Relapse prevention for alcohol and drug problems: That was Zen, this is Tao. Addictive behaviors: New readings on etiology, prevention, and treatment; 2009.
43. Hawker C, Merkouris S, Youssef G, Dowling N. Exploring the associations between gambling cravings, self-efficacy, and gambling episodes: An Ecological Momentary Assessment study. Addictive Behaviors 2021 Jan;112:106574. [doi: 10.1016/j.addbeh.2020.106574]
44. ACIL Allen Consulting, Deakin University, Central Queensland University, Social Research Centre. Fourth Social and Economic Impact Study of Gambling in Tasmania. Tasmania: Tasmanian Department of Treasury and Finance; 2017. URL: https://acilallen.com.au/projects/other/fourth-social-and-economic-impact-study-of-gambling-in-tasmania-volume-2 [accessed 2021-03-10]
45. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. Health Psychology 2015 Dec;34:1209-1219. [doi: 10.1037/hea0000306]
46. Attkisson C, Greenfield T. The client satisfaction questionnaire (CSQ) scales. Outcome assessment in clinical practice: Baltimore: Williams & Wilkins; 1995.
47. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: A New Tool for Assessing the Quality of Health Mobile Apps. JMIR mHealth uHealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422]
48. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and Analysis of Existing Mobile Phone Apps to Support Heart Failure Symptom Monitoring and Self-Care Management Using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74. [doi: 10.2196/mhealth.5882]
49. Carpenter SM, Menictas M, Nahum-Shani I, Wetter DW, Murphy SA. Developments in Mobile Health Just-in-Time Adaptive Interventions for Addiction Science. Current Addiction Reports 2020 Jun 27;7(3):280-290. [doi: 10.1007/s40429-020-00322-y]
50. Witkiewitz K, Desai SA, Bowen S, Leigh BC, Kirouac M, Larimer ME. Development and evaluation of a mobile intervention for heavy drinking and smoking among college students. Psychology of Addictive Behaviors 2014 Sep;28(3):639-650. [doi: 10.1037/a0034747]
51. Kim SW, Grant JE, Potenza MN, Blanco C, Hollander E. The Gambling Symptom Assessment Scale (G-SAS): A reliability and validity study. Psychiatry Research 2009 Mar;166(1):76-84. [doi: 10.1016/j.psychres.2007.11.008]
52. Ferris J, Wynne H. The Canadian Problem Gambling Index: Final report. Ottawa, ON: Canadian Centre on Substance Abuse; 2001.
53. StataCorp (Computer program). StataCorp. Version: College Station, TX: StataCorp LP; 2013.
54. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
55. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edn. Hillsdale, New Jersey: Lrlbaum; 1988:E.
56. Christensen L, Mendoza JL. A method of assessing change in a single subject: An alteration of the RC index. Behavior Therapy 1986 Jun;17(3):305-308. [doi: 10.1016/s0005-7894(86)80060-0]
57. Hageman WJ, Arrindell WA. A further refinement of the reliable change (RC) index by improving the pre-post difference score: Introducing RCID. Behaviour Research and Therapy 1993 Sep;31(7):693-700. [doi: 10.1016/0005-7967(93)90122-b]
58. Maust D, Cristancho M, Gray L, Rushing S, Tjoa C, Thase M. Psychiatric rating scales. Handbook of Clinical Neurology 2012;106:227-237. [doi: 10.1016/b978-0-444-52002-9.00013-9]
59. Babor T, Grant M. From clinical research to secondary prevention: International collaboration in the development of the Alcohol Disorders Identification Test (AUDIT). Alcohol Health & Research World 1989;13(4):371-375.
60. Ma X, Zhang J, Zhong W, Shu C, Wang F, Wen J, et al. The diagnostic role of a short screening tool—the distress thermometer: a meta-analysis. Supportive care in cancer 2014 Feb 8;22(7):1741-1755. [doi: 10.1007/s00520-014-2143-1]
61. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. Journal of Consulting and Clinical Psychology 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459]
62. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments. JMIR Mental Health 2016 Mar 01;3(1):e7. [doi: 10.2196/mental.4984]

63. Hing N, Russell AM, Thomas A, Jenkinson R. Hey Big Spender: An Ecological Momentary Assessment of Sports and Race Betting Expenditure by Gambler Characteristics. Journal of Gambling Issues 2019 Jun 20;42:42-61. [doi: 10.4309/jgi.2019.42.3]
64. Cowlishaw S, Merkouris S, Dowling N, Anderson C, Jackson A, Thomas S. Psychological therapies for pathological and problem gambling. Cochrane Database of Systematic Reviews 2012;11:1-88. [doi: 10.1002/14651858.cd008937.pub2]
65. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Translational Behavioral Medicine 2016 Dec 13;7(2):254-267. [doi: 10.1007/s13142-016-0453-1]
66. Hufford MR, Shields AL, Shiffman S, Paty J, Balabanis M. Reactivity to ecological momentary assessment: An example using undergraduate problem drinkers. Psychology of Addictive Behaviors 2002 Sep;16(3):205-211. [doi: 10.1037/0893-164x.16.3.205]
67. Attwood S, Parke H, Larsen J, Morton KL. BMC Public Health 2017 May 17;17(1):1-21. [doi: 10.1186/s12889-017-4358-9]
68. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS ONE 2017 Feb 2;12(2):1-20. [doi: 10.1371/journal.pone.0170581]

Abbreviations

CBT: cognitive behavioral therapy
CSQ-3: Client-Satisfaction Questionnaire-3
EMA: ecological momentary assessment
EMI: ecological momentary intervention
G-SAS: Gambling-Symptom Assessment Scale
MARS: Mobile App Rating Scale
MI: motivational interviewing
PGSI: Problem Gambling Severity Index
RCI: Reliable Change Index
VAS: Visual Analogue Scale

Edited by R Kukafka; submitted 16.11.20; peer-reviewed by H Gray, S Badawy; comments to author 23.12.20; revised version received 31.12.20; accepted 31.01.21; published 26.03.21.

Please cite as:
Hawker CO, Merkouris SS, Youssef GJ, Dowling NA
A Smartphone-Delivered Ecological Momentary Intervention for Problem Gambling (GamblingLess: Curb Your Urge): Single-Arm Acceptability and Feasibility Trial
J Med Internet Res 2021;23(3):e25786
URL: https://www.jmir.org/2021/3/e25786
doi:10.2196/25786
PMID:33769294

Original Paper

Health Care Students' Knowledge of and Attitudes, Beliefs, and Practices Toward the French COVID-19 App: Cross-sectional Questionnaire Study

Ilaria Montagni[1], DPhil; Nicolas Roussel[2], DPhil; Rodolphe Thiébaut[1,2,3*], DPhil, MD; Christophe Tzourio[1,3*], DPhil, MD

[1]Bordeaux Population Health Research Center, U1219, Bordeaux University, INSERM, Bordeaux, France

[2]Inria, Bordeaux University, Bordeaux, France

[3]Hospital Center Bordeaux University, Bordeaux, France

[*]these authors contributed equally

Corresponding Author:
Ilaria Montagni, DPhil
Bordeaux Population Health Research Center, U1219
Bordeaux University, INSERM
146 rue Léo Saignat
Bordeaux, 33000
France
Phone: 33 05 57 57 16 5
Email: ilaria.montagni@u-bordeaux.fr

Abstract

Background: Many countries worldwide have developed mobile phone apps capable of supporting instantaneous contact tracing to control the COVID-19 pandemic. In France, a few people have downloaded and are using the StopCovid contact tracing app. Students in the health domain are of particular concern in terms of app uptake. Exploring their use and opinions about the app can inform improvements and diffusion of StopCovid among young people.

Objective: The aim of this study is to investigate health care students' knowledge of and attitudes, beliefs, and practices (KABP) toward the StopCovid app.

Methods: A field survey was conducted among 318 students at the health sciences campus of the University of Bordeaux, France, between September 25 and October 16, 2020. A quota sampling method was used, and descriptive statistics and univariate analyses were performed.

Results: Of the 318 respondents, 77.3% (n=246) had heard about the app, but only 11.3% (n=36) had downloaded it, and 4.7% (n=15) were still using it at the time of the survey. Among the 210 participants who had heard about the app but did not download it, the main reasons for not using the app were a belief that it was not effective given its limited diffusion (n=37, 17.6%), a lack of interest (n=37, 17.6%), and distrust in the data security and fear of being geolocated (n=33, 15.7%). Among the 72 students who had not heard of the app and were given a brief description of its functioning and confidentiality policy, 52.7% (n=38) said they would use it. Participants reported that the main solution for increasing the use of the app would be better communication about it (227/318, 71.4%).

Conclusions: Even among health students, the contact tracing app was poorly used. The findings suggest that improved communication about its advantages and simplicity of use as well as clarifying false beliefs about it could help improve uptake.

(J Med Internet Res 2021;23(3):e26399) doi:10.2196/26399

KEYWORDS

contact tracing; COVID-19; mobile app; students; field survey; app; survey; monitoring; knowledge; attitude; belief; practice; communication; use

Introduction

Background Context

Nonpharmaceutical interventions have been used to contain the spread of the SARS-CoV-2 virus [1] while effective treatments and optimal vaccine coverage are made. Besides generalized lockdown and barrier gestures, one of the solutions to limit contagion, locate clusters, and isolate them is the tracing of infected people. Contact tracing is a systematic method used as part of a disease surveillance strategy (predict, observe, and minimize) [2]. In contact tracing, an index case with confirmed infection is asked to provide information about contacted people who were at risk of acquiring infection from the index case within a given time period (between 1 week and 14 days) before the positive test result. These contacts are then alleged to be tracked, advised about their risk, quarantined, and tested [3]. Conventional or manual contact tracing is a long process demanding human resources to contact and follow up with people one by one. It can engender several delays and is potentially biased by imperfect recall of contacts [4]. These limitations can be compensated by digital contact tracing [3,5].

Several smartphone apps have been developed worldwide to automatically and rapidly trace contacts in real time. Across all continents more than 45 apps are currently used [6] and several states are planning to launch such apps [7]. The general functioning of these apps is that each mobile device running the app keeps track of other mobile devices running the app that it comes in to close contact with. When users inform their app that they tested positive, this contact log is used to determine the other mobile devices—and users—that should be notified. Existing apps use different technologies and algorithmic methods to detect contacts between mobile devices (eg, short range Bluetooth Low Energy information exchange or GPS-, WiFi-, or Bluetooth-based geolocation), to keep track of these contacts (eg, using temporary unique identifiers), to evaluate the infection risk (eg, based on a predicted distance and the duration of the contact), and to notify potentially exposed people using a centralized or decentralized network approach [8]. The effectiveness of these apps is based on the fact that individuals are systematically tested, that results of these tests are correct and communicated in the app, that the individuals who are in contact have a smartphone, and that a high proportion of smartphone users download and use the app so as to interrupt the chains of infection transmission [9].

The recrudescence of the virus after the general lockdown from March to May 2020 has especially concerned young people across France and in the Bordeaux region in particular, where incidence of COVID-19 positive cases among young adults aged 20-30 years has increased to about 252/100,000 per week (weeks 41-42) [10]. Since September 1, 2020, several hundred students of the University of Bordeaux have been tested, and 26 of them returned positive results as of November 10. Students in the health domain are on the front line in terms of contagion. First, as many students across France, they are at risk because of contact with their peers. They often meet at the university (before, during, and after classes), downtown, or for private events. During these encounters, barrier gestures and preventive measures are not always respected. Second, their role as future health-related workers might suppose that they should set the example, since they are sensitized to adopt behaviors in favor of health promotion and prevention. Third, most students in the health domain are in contact with patients directly or indirectly through their interaction with health care workers. These different situations, informal, unprotected, and in relation with potentially unknown people, are typically those in which contact tracing apps make the most sense. Furthermore, students are digitally literate [11] and are supposed to be more at ease with the downloading and use of apps. French students in the health domain are thus a priority target for the uptake of the French COVID-19–related contact tracing app.

The StopCovid App

The contact tracing app “StopCovid” was launched by the French Government on June 2, 2020 [12]. It was developed by a team of public and private partners lead by the French National Institute for Research in Digital Science and Technology (Inria), and was available in both the Apple and Google Play stores, as it worked on iOS and Android phones. The app was based on Bluetooth signals running in the background of the phone with low-energy wireless transmission [13]. Once the app was activated, the phone logged other phones it came in to contact with, assuming these devices were running StopCovid. These logs did not include any identifying information about the user; they used random ID codes that changed every 15 minutes and were deleted once they were older than 14 days (the incubation period for COVID-19). The app did not locate the user (no GPS-, WiFi-, or Bluetooth-based geolocation); it only knew which random IDs the phone had come in to contact with. Being transparent and anonymous, the app did not collect any personal data nor contact details. If a user declared being positive for COVID-19 using a code delivered with the test results, the app would send that record of the rotating IDs to a centralized server, which in turn would send them out to other devices using the system [14]. Anyone that had the app activated who had been nearby in the last 2 weeks would be pinged with an alert. More precisely, this notification was sent if the person had spent more than 15 minutes within 1 meter of an index case. Users were then recommended to inform their general practitioner, get tested, and self-isolate, thus potentially stopping another line of transmission.

As of October 2020, StopCovid had been installed more than 2.7 million times since the beginning of June (about 4% of the French population, 67 million). Only 7969 users had declared being COVID-19 positive in the app, and only 472 notifications had been sent to potential at-risk contacts. The uptake was less than for apps in Germany (downloaded by 18 million people, about 21% of the German population, 84 million), England and Wales (16 million downloads in a population of 59 million, 27%), and Italy (9 million downloads in a population of 60 million, 26%). All these percentages are low considering that approximately 60% of the adult population would have to adopt the app to contain the pandemic [4,5]. In general, statistics show a limited use of these apps in Europe [15].

The effectiveness of the StopCovid app must be framed in the specific French context: the testing strategy was and is still

unclear with limited testing capacities, tracing was not always possible, and self-quarantining was voluntary and not always followed [16]. Furthermore, three sources of risk have been identified in the StopCovid app in terms of security and data protection: (1) the hacking of the central database, (2) the reporting of fictitious or unverified cases of infection, and (3) the increased vulnerability of the smartphones themselves caused by the activation of Bluetooth [7].

Literature Review on the Uptake of COVID-19–Related Contact Tracing Apps

Recently, several researchers have investigated the acceptability and use of contact tracing apps in the context of the COVID-19 pandemic. Some studies are based on surveys assessing the uptake of these apps among different population samples. These studies mostly refer to a hypothetical app and the intention to use it [16-24], and only a few collect information on the use of an existing tool like StopCovid [25-27]. The majority of documents reporting the real uptake of contact tracing apps are national statistics without a scientific and theoretical background. Other studies are critical viewpoints arguing on the ethical, technical, political, and scientific impact of contact tracing apps on society [2,7,9,28,29].

Concerning surveys, a multicountry cross-sectional study on 1849 adults across France, Germany, Italy, the United Kingdom, and the United States [17] showed that 74.8% of the respondents would install or keep a contact tracing app. Concerns about cybersecurity and privacy, together with a lack of trust in the government, were mentioned as the main barriers to app adoption. Another survey was conducted on 406 German adults [18], and the results showed that trust in the official app providers played an important role in the contact tracing app uptake. However, the threat appraisal of potential infection was not related to the motivation for using the app or for providing one's own infection status to it. In Belgium [19], 48.7% of 1500 adults declared intending to use a COVID-19 tracing app. The most important predictor was the perceived benefits of the app. Respondents also reported that the clarity on how the app functioned was correlated to the will to use it. Dutch citizens were interviewed in two studies [20,23]: 41.2%-64.1% of the respondents (n=238 [20] and n=900 [23], respectively) were willing to use a contact tracing app. In one study [20], the main reason to use such an app was to control the spread of COVID-19 (30.6%). Concerns about privacy were mentioned as the main reason for not using the app (64.8%). In the other study [23], the rate of potential users strongly varied by age group: the adoption rates of the app ranged from 45.6% to 79.4% for people in the oldest (≥75 years) and youngest (15-34 years) age groups. Educational attainment, the presence of serious underlying health conditions, and the respondents' stance on COVID-19 infection risks were also correlated with the predicted adoption rate. A national online survey on the Irish population (n=8088 responses) [21] showed that 84% of respondents would probably or definitely download the app. The most common reason for downloading the app was helping family members and friends (79%), and with a sense of responsibility to the wider community (78%). The most common reason for not downloading the app was fear that technology companies or the government might use the app technology for greater surveillance after the pandemic (41%). A longitudinal study was also conducted in Luxembourg on a representative sample of 730 adults [22]. The results showed that 72% would probably or definitely install the app if one was made available. Among motives in favor of contact tracing apps, respondents consistently mentioned responsibility toward the community and loved ones. In contrast, 11% of respondents would definitely not install the app, and their general willingness to use one was hampered by privacy and data security issues.

Acceptance of COVID-19 contact tracing apps has also been explored in France. In the first survey [27], 44% of a representative sample of 2000 French people declared that they would accept being electronically traced to avoid the spread of the virus. However, 23% were definitely against the app, and the majority of them were males and aged 25-34 years. The main reason for opposition was the fear of losing one's freedom. Another recent survey on 1849 French adults [16] showed that the contact tracing app was rather or totally acceptable by 42.1% of the respondents. A positive correlation was found between the perceived health consequences in case of COVID-19 infection and the willingness to use the contact tracing app. Trust in the government to handle the health crisis was also strongly and positively correlated with the potential use of the app.

Concerning critical viewpoints and opinion papers, they describe the public debate on privacy concerns due to the sensitive nature of the collected data. In particular, several researchers have argued that the adoption of contact tracing apps could lead to the economic exploitation of private data and might create a mass electronic surveillance system [7]. European governments have largely debated on the use of these apps, and ethical guidelines to develop and diffuse them have also been formulated [28]. In France, researchers have particularly investigated within a theoretical framework why the population has not largely adopted the StopCovid app [2,9]. According to an opinion paper [9], there are three main reasons for the low uptake of the app: the belief that the app will not be effective because we cannot reasonably expect that its adoption rate will be sufficient to be protective, the fear of data privacy breaches due to Bluetooth and to the centralized architecture of the app, and concerns on long-term surveillance and informational privacy. According to the author, the app raises a privacy paradox [30] where immediate benefits (eg, the reduction of contacts with infected people) are preferred to the value of privacy. Since the app does not seem to be effective given its limited use, it is not worth risking the breach of one's privacy. A second study [2] analyzed the political and scientific discourse around the promotion of the StopCovid app. Digital solutions like contact tracing apps might represent a form of alienation including government distrust. By collecting and analyzing media, scientific, and policy articles mentioning StopCovid, the study reported the contradictions of the government in handling the COVID-19 crisis based on partial and imprecise knowledge about the virus. In this context, government officials did not explain in plain language the security, privacy, data collection, processing, storage, and reuse of the StopCovid app. The app was then considered as not efficacious because of its low uptake,

characterized by lack of transparency and based on alienation and coercion.

Study Setting and Aim

This study was conducted at the beginning of the academic year at the University of Bordeaux, France, when face-to-face education was re-established, and students could freely circulate after the first general lockdown. The StopCovid app was available and downloadable for 4 months. The aim of this study is to describe knowledge, attitudes, beliefs, and practices (KABP) about the StopCovid app among students in the health domain in the Bordeaux region. The expected impact is informing on potential improvements as well as public-oriented communication strategies and appropriate political decisions to increase the app diffusion.

Methods

The Field Survey: Recruitment

This study was conducted within the framework of the larger ongoing i-Share (Internet-Based Students Health Research Enterprise [31]) cohort study, a French, nationwide web-based survey on the health and well-being of university students, whose principal investigators and operational staff are based at the University of Bordeaux [32].

The field survey consisted of a paper questionnaire administered face-to-face by five undergraduate students (interviewers) who had been trained to take notes, fill in the questionnaire, and describe the app to respondents. Interviewers approached their peers in the halls, canteen, courtyards, library, and study rooms at the health sciences campus of the University of Bordeaux. The collection of the data started on September 25, 2020, and ended on October 16, 2020. A sample size of 300 respondents was targeted with quotas set for the sample to be representative of the overall population of students in the health domain at the University of Bordeaux (n=16,566) in terms of sex, age (18-30 years), specific field of health-related study (medicine, dentistry, nursing, pharmacy, public health, etc), and year of study (1 to >6 years). The inclusion criteria were being older than 18 years, being a student in the health domain enrolled at the University of Bordeaux, and providing oral informed consent.

The Questionnaire

The questionnaire was co-designed and tested by a team of 14 public health researchers and operational staff following a structured survey construction method in five steps [33]. The final questionnaire was composed of 36 items, 14 of which were common to all students (sociodemographic characteristics, suggestions for increasing the diffusion of the app, willing to recommend the app to family and friends, and fake news about data collection and sharing within the app). The other items were administered based on four different scenarios: (1) the student has already heard about the app and has downloaded it, (2) the student has already heard about the app and has not downloaded it, (3) the student has never heard about the app but would download it, and (4) the student has never heard about the app and would not download it. Specific questions were then asked depending on the scenario. Before answering further questions, students who had not heard about the app were provided a brief description of it. After responding to fake news about data collection and sharing within the app, all students were given the correct answers. Some questions were multiple choice items. The English version of the questionnaire is available in Multimedia Appendix 1. The time of administration and completion of the questionnaire was about 10 minutes. The field survey was approved by the University of Bordeaux. The oral informed consent reassured students of the anonymous format of the survey and that use of collected data was for research purposes only.

Theoretical Framework

The questionnaire was based on the KABP scheme, which stands for the assessment of knowledge, attitudes, beliefs, and practices of populations about a specific health-related topic. This scheme is extensively used as a quantitative method (predefined questions formatted in a standardized questionnaire) that provides access to quantitative and qualitative information. Thus, items of the questionnaire were developed to capture students' KABP about the StopCovid app. The results are discussed following the four components of this scheme.

Collected data were interpreted a posteriori through the prism of the technology acceptance model (TAM) [34] and the protection motivation theory (PTM) [35]. The TAM posits that an individual's intent to use (ie, accept) a technology and use behavior (ie, actual use) is influenced by perceived ease of use and usefulness, which are mediated by external variables such as individual differences, system characteristics and complexity, and social influences. The TAM is especially adaptable to technology-related motivations. The PTM explains why people adopt a preventive behavior and what role fear appeals play in this process. This model comprises the threat appraisal of a potential risk (eg, infection with SARS-CoV-2) and coping appraisal of the recommended preventive behavior (eg, using StopCovid) [18]. Threat appraisal includes the perceived severity of the disease and vulnerability to it. Coping appraisal includes perceived self-efficacy (ie, belief in one's own competence to perform a behavior even in the face of barriers) and response efficacy (ie, individuals are convinced that a behavior leads to the desired outcome and will be more likely to intend to perform the behavior). The PTM is adaptable to both health-related and technology-related motivations [18].

Data Analysis

All data from the paper questionnaires were entered by the student interviewers in a digital database through the EpiData software version 3.1 [36]. A descriptive analysis was performed, presenting all variables and measures in the form of numbers and percentages for qualitative variables and means and SDs for quantitative variables. Chi-square or exact Fisher frequency comparison tests were used to identify statistically significant differences by age, gender, field, and year of study, modified to binary variables if necessary. Data were normally distributed. Statistical significance was defined with a P value <.05. Statistical powers were calculated for each frequency comparison test, chi-square or Fisher exact test, with the condition of a minimum sample size for the Fisher exact test. In calculating the power, an approximation of the normal distribution for the chi-square tests or an approximation of the

Walters normal distribution for the Fisher exact tests was used. In addition, the size and proportion for each group was specified and the alpha was set at .05. The data were analyzed with SAS version 9.3 (SAS Institute).

Results

A total of 590 students were approached to complete the survey after a brief explanation of its objective; 318 completed the survey, while 272 refused to participate, creating a final participation rate of 53.9%. Reasons for not participating were lack of time or no interest in the study topic. Of the 318 participants, 65.7% (n=209) were female students, and the mean age was 20.4 (SD 2.39) years. All fields and years of study were represented. The majority (n=193, 60.7%) of the participants were medical students, which is in line with the total number of medical students at the University of Bordeaux. In accordance with university-related statistics, first-year students were also the most represented (n=129, 40.6%). Figure 1 shows the flowchart of the study population, and Table 1 shows the sociodemographic characteristics with the corresponding data for the total population of health-related students at the University of Bordeaux.

Figure 1. Flowchart of the study population (n=318).

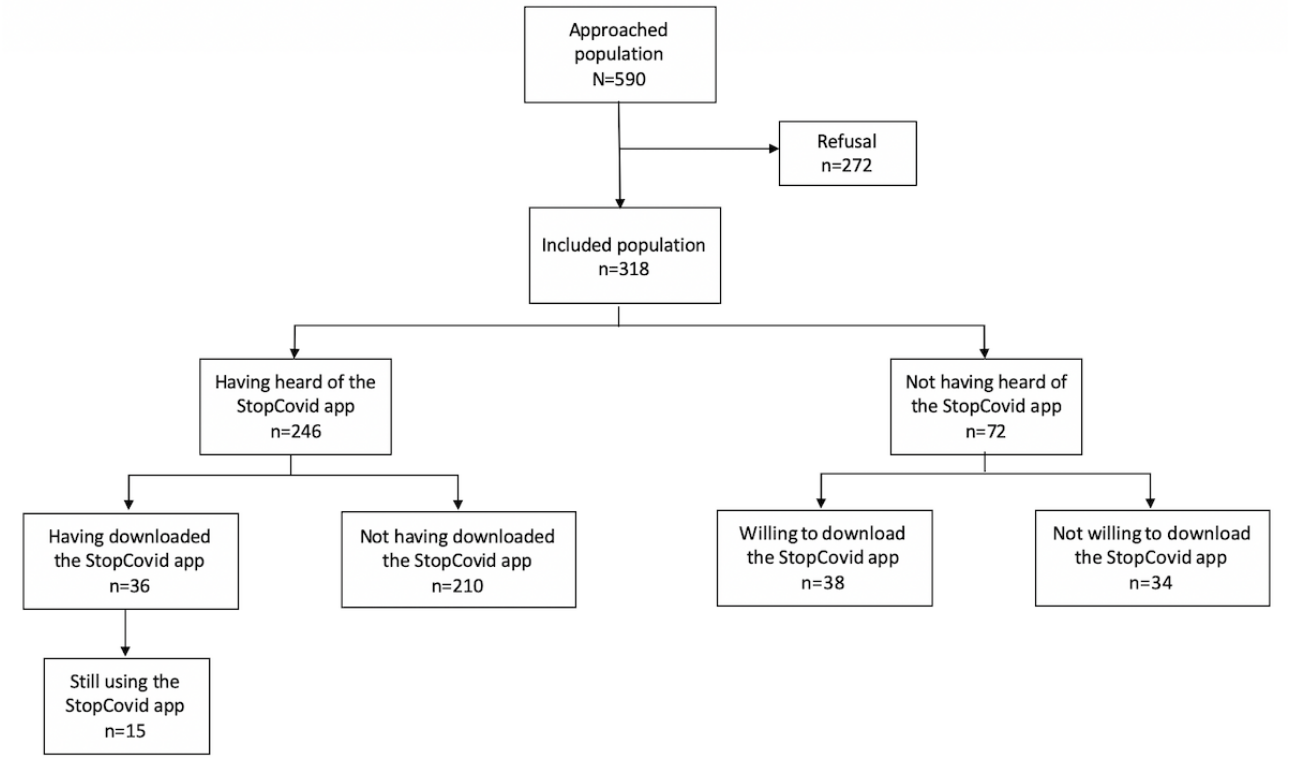


Table 1. Sociodemographic characteristics of the study population and comparison with all students in the health domain.

Characteristics	Study population (n=318)	Total health care student population at the University of Bordeaux (n=16,566)[a]
Sex, n (%)		
Female	209 (65.7)	11,713 (70.7)
Male	109 (34.3)	4853 (29.3)
Age (years), mean (SD)	20.4 (2.39)	23.8 (—[b])
Year of study, n (%)		
1	129 (40.6)	4445 (31.2)
2	64 (20.1)	2341 (16.4)
3	46 (14.5)	2334 (16.4)
4	23 (7.2)	1200 (8.4)
5	44 (13.8)	1267 (8.9)
>5	10 (3.1)	2651 (18.6)
Other	2 (0.6)	—
Field of study, n (%)		
Medicine	193 (60.7)	7104 (49.9)
Pharmacy	61 (19.2)	1138 (8.0)
Dentistry	12 (3.8)	521 (3.7)
Nursing	14 (4.4)	3972 (27.9)
Public health	15 (4.7)	—
Other	23 (7.2)	1503 (10.6)

[a]Data obtained from internal university documents.

[b]Data not available.

The majority (n=246, 77.3%) of the participants had already heard about the app, mostly through the media (216/246, 87.8%) and secondly through family and friends (39/246, 15.9%). Concerning these variables, no statistically significant differences were found based on age (*P*=.09), sex (*P*=.85), field (*P*=.08), or year of study (*P*=.06).

Most of the 246 students that knew of the app correctly knew that the app was promoted by the government (n=179, 72.8%), but 25.2% (n=62) answered that they did not know who the promoter was. Male students knew significantly more than female students that the app was promoted by the government (69/85, 81.2% vs 110/161, 68.3%; *P*=.03). Female students were significantly more likely to ignore the promoter of the app compared to male students (47/161, 29.2% vs 15/85, 17.6%; *P*=.047). Students of any health-related discipline other than medicine responded significantly more than medical students that the app was promoted by a research laboratory (4/103, 3.9% vs 0/143, 0.0%; *P*=.03). Medical students were significantly more likely to ignore the promoter of the app (43/143, 30.1% vs 19/103, 18.4% ; *P*=.04). No statistically significant differences were found based on age (*P*=.06) or year of study (*P*=.17).

Among the 246 participants who had heard about the app, 14.6% (n=36) had actually downloaded it when it was first released in June 2020 (22/36, 61.1%) or with the new cases of COVID-19 at the beginning of the university year (6/36, 16.7%). Of these 36 students, 41.6% (n=15) were still using the app. Of the total 318 participants, 4.7% (n=15) of students were using the app at the moment of the survey, while those who uninstalled the app had used it from 1 day (6/36, 16.7%) to several weeks (6/36, 16.7%). The main reasons for uninstalling the app were that it was not useful (14/21, 66.7%), the respondent forgot to activate the Bluetooth (5/21, 23.8%), the app drained the phone battery (4/21, 19.0%), and too few people were using it thus making the app ineffective (4/21, 19.0%). Accordingly, students reported that the main fault of the app was that it seemed inefficient given its limited uptake (17/35, 48.6%; 1 missing). For 25.7% (9/35), the app presented technical problems like draining the battery, depending on Bluetooth, or occupying too much storage on the phone. Concerning all previous variables, no statistically significant differences were found based on age (*P* values ranging from .27 to >.99), sex (*P* values ranging from .19 to >.99), field (*P* values ranging from .13 to >.99), or year of study (*P* values ranging from .26 to >.99).

Some of these students also reported that its qualities were that it was easy to use (18/35, 51.4%) and that it was reassuring (9/35, 25.7%). Male students found the app significantly more user-friendly than female students (12/17, 70.6% vs 6/18, 33.3%; *P*=.02). Concerning this variable on the quality of the app, no statistically significant differences were found based on age (*P*=.39), field (*P*=.33), or year of study (*P*=.18).

Reasons for downloading or not downloading the app are shown in Tables 2 and 3 (multiple answers possible for each individual).

Among the 210 participants who had heard about the app but did not download it, the main reasons for not using the app were lack of interest (n=90, 42.9%), belief that it was neither effective nor useful given its limited diffusion (n=37, 17.6%), not having time to think about it (n=37, 17.6%), and distrust in data security and fear of being geolocated (n=33, 15.7%). The majority of these students might change their mind and use the app if they had more information about it through better communication strategies (n=61, 29.0%) and if more people would use it (n=54, 25.7%). Nonetheless, 26.2% (n=55) would not change their mind and would still not download the app. On the other hand, the main reasons for downloading the app were out of curiosity (13/36, 36.1%) and to protect one's family, others, and oneself from possible infection (13/36, 36.1%).

Table 2. Reasons for downloading the StopCovid app.

Reasons[a]	Yes, I have downloaded the app (n=36), n (%)	Yes, I would download the app (n=38), n (%)
Reasons for downloading the app		
Out of curiosity	13 (36.1)	14 (36.8)
To protect my family, others, and myself from possible infection	13 (36.1)	24 (63.2)
Because the government advised downloading of the app	5 (13.9)	0 (0.0)
The app could be useful to contain the spread of the virus in general	5 (13.9)	20 (52.6)
I am afraid of the virus, and all strategies are good to avoid it	1 (2.8)	1 (2.6)
I was reassured the app was anonymous	N/A[b]	0 (0.0)
Other	11 (30.6)	1 (2.6)

[a]Multiple answers possible.

[b]N/A: not applicable.

Table 3. Reasons for not downloading the StopCovid app.

Reasons[a]	No, I have not downloaded the app (n=210), n (%)	No, I would not download the app (n=34), n (%)
Reasons for not downloading the app		
Cannot see the interest or need	90 (42.9)	17 (50.0)
Do not like the general idea of this app	10 (4.8)	1 (2.9)
Do not know how it works, did not get enough information on the app	23 (11.0)	2 (5.9)
I am suspicious of this type of app	18 (8.6)	1 (2.9)
Do not trust, because I do not know who is offering this app	11 (5.2)	1 (2.9)
Not sure about the security of the data, fear of geolocation	33 (15.7)	1 (2.9)
My family and friends have discouraged me from downloading it	1 (0.5)	0 (0.0)
No storage on my phone or it is not powerful enough to have an extra app (battery, Bluetooth)	27 (12.9)	6 (17.6)
Do not carry my phone with me at all times	2 (1.0)	4 (11.8)
Do not use public transportation and/or do not go out much in public places (do not come into contact with strangers)	11 (5.2)	4 (11.8)
It does not seem to be effective (too few people use it)	37 (17.6)	8 (23.5)
Heard negative feedback on this app	7 (3.3)	1 (20.9)
Do not really have time to think about it	37 (17.6)	3 (8.8)
By negligence, not concerned	28 (13.3)	5 (14.7)
Not sure what it is all about, the principle and/or the functioning	11 (5.2)	N/A[b]
Other	15 (7.1)	2 (5.9)

[a]Multiple answers possible.

[b]N/A: not applicable.

The 72 students who had never heard about the app were asked to imagine its content and objective: 41.7% (n=30) reported that it was an app providing advice and information about COVID-19, 29.2% (n=21) reported that it was an app to limit the spread of the virus, 29.2% (n=21) did not know, and 15.3% (n=11) answered "other." After a short description of the app, 52.7% (n=38) said they would download it. The reasons for downloading or not downloading the app are similar to those provided by the sample who had heard about the app. Among the 34 students who had never heard about the app and were still not willing to download it after a brief description, 32.4% (n=11) would not change their mind, 17.6% (n=6) would download it if more people used it, and 11.8% (n=4) would download it if they had a better mobile phone.

Concerning the functioning of the app, 83.3% (30/36) of the respondents said that they were able to explain it. However, when further asked about geolocation, access to contact information, and how data were transmitted and stocked, their answers were mostly incorrect. As expected, students who had not heard about the app before, but who were presented a quick description of it during the survey, provided correct answers more than their peers. Detailed results are shown in Table 4.

Table 4. Knowledge and beliefs about the functioning and data management of the StopCovid app.

Knowledge and beliefs	Yes, I have heard about the app (n=246), n (%)	No, I have not heard about the app (n=72), n (%)	Total (N=318), n (%)
StopCovid geolocates you and tracks your movements			
No (correct answer)	78 (31.7)	44 (61.1)	122 (38.4)
Yes	138 (56.1)	19 (26.4)	157 (49.4)
Not sure	30 (12.2)	9 (12.5)	39 (12.3)
StopCovid collects your contacts and knows their names (on the phone, on social networks, etc)			
No (correct answer)	178 (72.4)	56 (77.8)	234 (73.6)
Yes	37 (15.0)	5 (6.9)	42 (13.2)
Not sure	31 (12.6)	11 (15.3)	42 (13.2)
StopCovid detects people around you and knows their names (physical contacts)			
No (correct answer)	124 (50.4)	45 (62.5)	169 (53.1)
Yes	82 (33.3)	15 (20.8)	97 (30.5)
Not sure	40 (16.3)	12 (16.7)	52 (16.4)
StopCovid has access to your personal data and communicates them			
No (correct answer)	191 (77.6)	55 (76.4)	246 (77.4)
Yes	28 (11.4)	5 (6.9)	33 (10.4)
Not sure	27 (11.0)	12 (16.7)	39 (12.3)

Finally, all 318 participants were asked about factors for increasing the use of the app. For the majority (n=227, 71.4%), the solution was a better communication strategy. Other factors were making the app compulsory (n=45, 14.2%), registering more COVID-19 cases (n=30, 9.4%), more information and explanations about the app (n=21, 6.6%), better technical features (n=10, 3.1%), and "other" (n=64, 20.1%).

Discussion

Principal Findings and Interpretation

As far as knowledge is concerned, 1 out of 5 students had never heard about the StopCovid app; this rate is surprisingly high considering that students in the health domain should be informed of existing tools to limit the spread of COVID-19. Those who knew the app had heard about it mostly through the media (216/246, 87.8%). The majority of students (179/246, 72.8%) correctly knew that the app was promoted by the government. However, concerning the functioning of the app, some students did not know how the contact tracing system worked and how data was managed: percentages of errors in describing the app ranged from 10.4% (33/318) to 49.4% (157/318). For them, the app was not straightforward; in the light of the TAM, reduced ease of app use determines lower acceptance. Furthermore, as an external variable, system complexity might have mediated the perception of ease of use for the app. In general, limited information about a tool, from knowing that it exists to knowing how it works, is associated with poor use of the tool itself. Consistently, when asked how they would improve StopCovid adoption, 71.4% (227/318) of students suggested deploying better communication and information strategies for increasing knowledge about the app.

In terms of attitudes, students reported several reasons for not downloading or uninstalling the app. Their intention not to use the app was mostly due to the fact that they considered the app as neither useful nor effective (14/19, 73.7%), especially because

few people were using it. As suggested by the TAM, perceived usefulness is a key determinant of acceptance of a new technology, which justifies the low adoption of StopCovid by students of our sample. Technical issues like draining the battery, use of Bluetooth, and mobile phone storage were also mentioned (9/19, 31.0%). Once again, according to the TAM, technological components are strictly related to acceptance of a digital tool. On the other hand, reasons to download the app included wanting to protect one's family and friends: percentages ranging from 36.1% (13/36) to 63.2% (24/38). According to the PMT, the effort or cost (ie, response efficacy) of using the app was worth it to protect others from the virus. A few students (5/36, 13.9%) reported that the promotion of the app by the government motivated them to use it. This result might reflect a certain degree of confidence in political authorities.

As for beliefs, when specifically asked about the functioning and data management of the app, half of the total sample believed that the app was intrusive: it could geolocate them, track their movements, and access phone contacts. For 1 out of 6 students, fear of being tracked and that data could be collected and shared discouraged them from downloading the app. This false belief might have been nourished by the fact that students could have heard in the media that data breaches were possible, directly on their phones through Bluetooth, and that central servers could be violated. According to the PMT model, if beliefs do not support the recommended preventive behavior, probability of adopting such behavior is reduced. Furthermore, if a data breach is felt like a threat, individuals would be motivated not to download the app to protect themselves. Among those who received a clear explanation of the functioning of the app and its confidentiality policy (no geolocation and no access to personal data), 1 out of 2 students felt reassured and would finally download the app. These results confirm that beliefs, either true or false, influence behavioral intention.

Finally, concerning practices, 14.6% (36/246) of participants had actually downloaded the app, and in the whole sample, only about 4.7% (15/318) were still using it at the time of the survey, which is in line with national statistics concerning the general French population (4%). Furthermore, 26.2% (55/210) of respondents would not change their mind and still would not use the app. Possible justifications could include the fact that young people might perceive the pandemic as not dangerous for them. Epidemiological data confirm that COVID-19 is fatal mostly for people older than 60 years or who have a chronic disease [37]. Within the PMT framework, considering the illness as not too severe and perceiving vulnerability as low are related to the limited need to adopt a specific health-related behavior, which corresponds, in our case, to the downloading of the app. Along the same line, students might feel their competences (ie, self-efficacy), such as barrier gestures, are enough to prevent the virus, independent of app use. These are potential explanations for the general lack of interest in the app showed by our sample (90/210, 42.9%).

Comparison With Prior Work

Overall, international and French surveys (eg, [16-18,27]) have showed a higher acceptance of a contact tracing app than the real use we found in our study. Percentages of potential use of the app range from 38.4% [16] to 84% [21], which are substantially higher than the 4.7% (15/138) of respondents in our study who were using StopCovid. However, the lowest rates of acceptance for a contact tracing app were found mostly in France: 38.4% [16] and 44% [27]. Inversely, in our sample, 26.2% (55/210; having heard about the app) to 32.4% (11/34; not having heard about the app) of students would not change their mind and would not use the app at all. This percentage is higher than in the Luxembourg survey (11%) [22] and the Irish survey (7%) [21], but similar to one of the French surveys (23%) [27] and the Belgian survey (20.4%) [19]. Discrepancies between our study and previous surveys might be explained by the fact that the latter asked hypothetical questions about future behavior; high levels of intended installations might not directly translate into actual installation. It might be harder for respondents to visualize how such apps work, thus limiting the reliability of their responses compared to a real-life scenario. Furthermore, optimistic results found in previous surveys might be due to the fact that they had been conducted when the epidemic was on the rise and before digital contact tracing had been widely discussed in the media, especially in relation to data security. This might be the case especially for France [16]; citizens' opinions might have changed when the StopCovid app was developed and controversies about it were raised in public debate in Spring 2020.

A study exploring the real uptake of an existing app in Singapore, the TraceTogether app, had an uptake of 20% [25]. This higher percentage, compared to our study, might be justified by the fact that Asian countries are often referred to for their decisive and authoritative responses to pandemics. More convincing communication around the app might have increased its uptake. Furthermore, TraceTogether has been a real pioneer in COVID-19–related apps given its high performance, which might have further supported its use. However, the TraceTogether app received criticism for draining mobile phone batteries, which was one of the faults reported in this study about the StopCovid app. In fact, excessive use of battery and data storage were mentioned by some of our students as reasons for not downloading the app (27/210, 12.9%), and 25% (9/36) had uninstalled the app because of these technical problems.

In our study, the three main reasons for not downloading the app were lack of interest (90/210, 42.9%), belief that it was neither effective nor useful given its limited diffusion (37/210, 17.6%), and not having time to think about it (37/210, 17.6%). No previous survey has reported these same reasons, even if in the literature the notion of contact tracing apps' effectiveness has been widely discussed [5]. Students were aware that if the app is not used by a consistent number of people, it is not efficacious at all. In general, our sample expressed disinterest in the app. The reasons should be further explored, but we might suppose that the app was not considered as useful given the other restrictive measures in place: national lockdown, social distancing, and barrier gestures.

Distrust in data security and fear of being geolocated were mentioned by our sample as the fourth reason for not downloading the app (33/210, 15.7%). Researchers worldwide, from Europe to Asia, have emphasized the privacy controversies of contact tracing apps, presenting them as the main fault of this type of technology [9,17,22,25,38]. Fears of greater surveillance and that the app might be hacked are mentioned in these studies as barriers to app use. In France, the question of data privacy related to the StopCovid app has been particularly explored; the app does not come without short-term and long-term risks of privacy and surveillance. French people face a moral dilemma: the app can prevent the spread of the pandemic, especially protecting older adults, but limit freedom to move, data security, and privacy, which are usually sensitive issues in the French culture and politics [9]. However, this did not seem to be a source of much concern in our study compared to perceived ineffectiveness and inutility of the app. For our sample, uptake of the app might not necessarily be a matter of data security or trust in the government but a question of practicality and usefulness. This result might be explained by the fact that the youth are already used to sharing their information online (eg, in social networks) and are not as concerned by cybersecurity [9]. In line with this, none of our respondents mentioned being reassured that it was anonymous as a reason to download the app. Furthermore, the proportion of students who received an explanation of the functioning of the app were comforted about the fact that no private data was collected, users were not geolocated, the app did not access contacts, and that only an anonymous code was transmitted to a centralized server by Bluetooth and deleted after 14 days. For the 15.7% (33/210) of students who were cautious about data security, following the PMT model, severity of and vulnerability to data misuse might have reduced their motivation to use the app, as reported in the German survey [18]. In any case, information should be more accurate on data security since this issue could discourage young people from downloading the app. Exact data management in the contact tracing app needs to be clarified to guarantee the respect of the user's privacy.

Similarly, trust in the authorities was mentioned in previous research as a factor influencing the uptake of the app: individuals who have less trust in their national government were also less supportive [17]. Despite not exploring the notion of trust in the government, we observed that 13.9% (5/36) of those who downloaded the app were motivated from the advice by the government. This response option might be considered as a proxy for trust in the government. Although data from a larger sample is needed to corroborate this result, we might assume that the political discourse has an impact on the diffusion of the app, whether positive or negative.

The main reasons for downloading the app were curiosity (13/36, 36.1% of those who downloaded the app and 14/38, 36.8% of those who would download the app) and wanting to protect family, others, and oneself from possible infection (13/36, 36.1% and 24/38, 63.2%, respectively). The second reason was also reported in the Luxembourg survey [22] and in the Irish study [21]. Students who had received an explanation of the functioning of the app reported twice as much as the other students the fact that the app could prevent them and their beloved ones from the spread of the virus. Similar to half of the Luxembourg study's sample [22], for some students, a good reason for installing the app was that it may stop the epidemic: percentages ranging from 13.9% (5/36) for those who downloaded the app to 52.6% (20/38) for those who would download the app. In general, when provided with a clear explanation of the app, 1 out of 2 students was convinced to download the app because of our study. In this line, 10.9% (23/210; of those who had heard about the app) and 5.9% (2/34; of those who had not heard about the app) of our respondents said that they would not download the app because they did not have enough information on how it worked. Similarly, participants in the survey conducted in Belgium [19] declared that lack of clarity on its functioning was among the reasons for not downloading the app. This suggests that providing clear information on the objectives of the app might promote its uptake.

Finally, we must consider the specificity of our population compared to previous research. Although studies were conducted on the general population (mostly nationally representative samples), we presented data exclusively from university students in the health domain. Our sample might have felt less concerned by the pandemic or simply less interested in this type of digital solution, or studying in the health domain and potentially working in hospitals might have made a contact tracing app for our respondents superfluous since they could be in contact with patients who were infected. A qualitative study would be useful to analyze the motivation for not using the app in this specific population.

Strengths and Limitations

This was one of the first studies reporting data on students' KABP about a contact tracing app in a pandemic context. Previous studies have explored the intention of downloading this type of app as a general idea but were not based on a developed and currently diffused app [17-19,23]. Reasons for downloading and using the app were presented to inform future steps to increase its diffusion. The specific focus on students was another strong point of this study: young people were especially concerned by the transmission of the virus in subsequent COVID-19 waves. Mobilizing this population to adopt the app is pivotal in this particular epidemiological context.

Limitations of this study include the relatively small sample. More than 300 students in the health domain were interviewed among a total population of 18,000 students. Findings cannot be generalized, but the sample was recruited according to quota sampling to be, as much as possible, representative of sex, age, specific field of study (from medicine to pharmacy), and year of study. This might increase the representativeness of the interviewed population group. However, it is possible that students interested in the topic were more willing to participate so that the final sample might be biased (self-selection bias). The small sample also justifies the few significant differences that were identified. This is confirmed by the low statistical powers that were obtained following performed statistical tests (<0.50).

The New Version of the StopCovid App: TousAntiCovid

StopCovid received several criticisms and even the French Prime Minister Jean Castex officially declared not having downloaded the app. The government considered the low uptake of the app as the main issue of StopCovid. Some weeks after the implementation of this study, on October 22, 2020, the French President Emmanuel Macron announced the launch of a new contact tracing app, TousAntiCovid. There are two main differences between the two versions of the app: embedded functionalities and promoting strategy. Concerning functionalities, they include provision of information on new cases (effective R, incidence rate, hospitalizations, etc), advice and news about COVID-19, geolocation of testing centers, and generation of the mandatory certificates for permission to be outdoors during the lockdown. This last functionality, in particular, might have increased the download and use of the app. As for promoting strategy, it was more intense but less coercive and more transparent about the app compared to StopCovid. The French President and Prime Minister were strongly engaged in the communication campaign from the beginning, whereas StopCovid had mainly been promoted by the Minister of Health and by the Digital Secretary of State [2], who have less influence on the general population. In light of this, a new survey on the TousAntiCovid app might provide different results to compare with our study.

Implications

This survey was conducted as the preliminary phase of a complex intervention aimed at promoting the uptake of the StopCovid app among students in the health domain at the University of Bordeaux. After this first appraisal of KABP about StopCovid, the next steps are to implement a series of actions at the university. Professors and lecturers have been mobilized and trained to present the contact tracing app to their students during classes. Furthermore, students will also be informed by more communication such as short videos on the university website and intranet, flyers, posts on social networks, and posters. Student ambassadors and associations will also be involved in the diffusion of the app. This complex intervention will be evaluated through a second series of random field surveys aimed at observing an increase in the number of app downloads. Depending on the results of the evaluation, the intervention will be extended to students in other fields of study at the University of Bordeaux and other universities across France.

Conclusion

Overall, we found broad support for app-based contact tracing, notwithstanding the low uptake of StopCovid among French students in the health domain. The results suggest that the functioning and purpose of the app were not well known and appraised among participants, especially because of the lack of factual communication. Efforts are to be taken in these terms to increase knowledge about the new TousAntiCovid app, diffuse its adoption, and consequently improve preventive behaviors among young people who represent an important target audience in the strategies to limit the transmission of COVID-19. The way the app traces contacts should be better explained so as to maximize its download and consequential use by eliminating any potential false belief. The French government should be particularly involved in providing quality, clear, appropriate, and straightforward information about TousAntiCovid.

Acknowledgments

The project is led by the University of Bordeaux in close collaboration with the University Hospital of Bordeaux, the Regional Health Authority, and the National Institute for Research in Digital Science and Technology (Inria).

The authors wish to thank the i-Share team [31] for managing and conducting the study and for revising the questionnaire: Julie Arsandaux, Fadi El Khoury, Shérazade Kinouani, Mélissa Macalli, Elena Milesi, Marie Mougin, Garance Perret, and Clothilde Pollet. Within the i-Share team, the authors wish to thank especially Edwige Pereira and Aude Pouymayou for their support in statistical analysis. The authors are also indebted to the student ambassadors that carried out the field survey: Idriss Boukili, Justine Char, Jade Giacosa, Oleg Hounkpatin, Jeanne Langlois, and Marie Lanneluc.

All data generated or analyzed during this study are included in this manuscript. The full data set is available on reasonable request.

The i-Share team is currently supported by an unrestricted grant of the Nouvelle-Aquitaine Regional Council (Conseil Régional Nouvelle-Aquitaine; grant no. 4370420) and by the Bordeaux *Initiatives d'excellence* program of the University of Bordeaux (ANR-10-IDEX-03-02). The team has also received grants from Public Health France (Santé Publique France, contract no. 19DPPP023-0) and the Nouvelle-Aquitaine Regional Health Agency (Agence Régionale de Santé Nouvelle-Aquitaine). The funding bodies were not involved in the study design or in the collection, analysis, or interpretation of the data.

Authors' Contributions

IM was responsible for writing the manuscript, conducting the literature search, data collection, and interpretation. CT, RT, and NR were responsible for the study design and revision of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
English version of the StopCovid field survey questionnaire.
[PDF File (Adobe PDF File), 63 KB - jmir_v23i3e26399_app1.pdf]

References

1. Kraemer MUG, Yang C, Gutierrez B, Wu C, Klein B, Pigott DM, Open COVID-19 Data Working Group, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 2020 May 01;368(6490):493-497 [FREE Full text] [doi: 10.1126/science.abb4218] [Medline: 32213647]
2. Rowe F, Ngwenyama O, Richet J. Contact-tracing apps and alienation in the age of COVID-19. Eur J Inf Syst 2020 Sep 13;29(5):545-562. [doi: 10.1080/0960085x.2020.1803155]
3. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. Lancet Public Health 2020 Aug;5(8):e452-e459 [FREE Full text] [doi: 10.1016/S2468-2667(20)30157-2] [Medline: 32682487]
4. Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. Lancet Digit Health 2020 Nov;2(11):e607-e621 [FREE Full text] [doi: 10.1016/S2589-7500(20)30184-9] [Medline: 32839755]
5. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):eabb6936 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
6. LI J, Guo X. Global deployment mappings and challenges of contact-tracing apps for COVID-19. SSRN J. Preprint posted online May 26, 2020. [doi: 10.2139/ssrn.3609516]
7. Jacob S, Lawarée J. The adoption of contact tracing applications of COVID-19 by European governments. Policy Design Pract 2020 Nov 28:1-15. [doi: 10.1080/25741292.2020.1850404]
8. John Leon Singh H, Couch D, Yap K. Mobile health apps that help with COVID-19 management: scoping review. JMIR Nurs 2020;3(1):e20596 [FREE Full text] [doi: 10.2196/20596] [Medline: 32897271]
9. Rowe F. Contact tracing apps and values dilemmas: a privacy paradox in a neo-liberal world. Int J Inf Manage 2020 Dec;55:102178 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102178] [Medline: 32836636]
10. Coronavirus?: chiffres clés et évolution de la COVID-19 en France et dans le Monde. Santé Publique France. 2020. URL: https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/coronavirus-chiffres-cles-et-evolution-de-la-covid-19-en-france-et-dans-le-monde [accessed 2021-02-26]
11. Shopova T. Digital literacy of students and its improvement at the university. Eries J 2014 Jul 31;7(2):26-32. [doi: 10.7160/eriesj.2014.070201]
12. StopCovid - TousAntiCovid. Government of France. 2020. URL: https://www.economie.gouv.fr/stopcovid [accessed 2021-02-26]
13. Cunche M, Boutet A, Castelluccia C, Lauradoux C. On using Bluetooth-low energy for contact tracing. Lyon 2020 Sep 01:2020.
14. Gorce J, Egan M, Gribonval R. An efficient algorithm to estimate Covid-19 infectiousness risk from BLE-RSSI measurements. Lyon 2020 May 28:1-40.
15. Contact tracing apps. Council of Europe. 2020. URL: https://www.coe.int/en/web/data-protection/contact-tracing-apps [accessed 2021-02-26]
16. Guillon M, Kergall P. Attitudes and opinions on quarantine and support for a contact-tracing application in France during the COVID-19 outbreak. Public Health 2020 Nov;188:21-31 [FREE Full text] [doi: 10.1016/j.puhe.2020.08.026] [Medline: 33059232]
17. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: cross-country survey study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
18. Kaspar K. Motivations for social distancing and app use as complementary measures to combat the COVID-19 pandemic: quantitative survey study. J Med Internet Res 2020 Aug 27;22(8):e21613 [FREE Full text] [doi: 10.2196/21613] [Medline: 32759100]
19. Walrave M, Waeterloos C, Ponnet K. Adoption of a contact tracing app for containing COVID-19: a health belief model approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
20. Jansen-Kosterink S, Hurmuz M, den Ouden M, van Velsen L. Predictors to use mobile apps for monitoring COVID-19 symptoms and contact tracing: a survey among Dutch citizens. medRxiv. Preprint posted online June 2, 2020. [doi: 10.1101/2020.06.02.20113423]
21. O'Callaghan ME, Buckley J, Fitzgerald B, Johnson K, Laffey J, McNicholas B, et al. A national survey of attitudes to COVID-19 digital contact tracing in the Republic of Ireland. Ir J Med Sci 2020 Oct 16:1-25 [FREE Full text] [doi: 10.1007/s11845-020-02389-y] [Medline: 33063226]
22. Riillo C, Peroni C, Sarracino F. Determinants of acceptability of contact tracing apps for COVID-19: initial results from Luxembourg. Economie et Statistiques 2020 Dec:1-31.

23. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mölken M. COVID-19 contact tracing apps: predicted uptake in the Netherlands based on a discrete choice experiment. JMIR Mhealth Uhealth 2020 Oct 09;8(10):e20741 [FREE Full text] [doi: 10.2196/20741] [Medline: 32795998]
24. Wiertz C, Banerjee A, Acar OA, Ghosh A. Predicted adoption rates of contact tracing app configurations - insights from a choice-based conjoint study with a representative sample of the UK population. SSRN J 2020 Apr 30:1-19 [FREE Full text] [doi: 10.2139/ssrn.3589199]
25. Goggin G. COVID-19 apps in Singapore and Australia: reimagining healthy nations with digital technology. Media Int Aust 2020 Aug 14;177(1):61-75. [doi: 10.1177/1329878x20949770]
26. Salathé M, Althaus C, Anderegg N, Antonioli D, Ballouz T, Bugnon E, et al. Early evidence of effectiveness of digital contact tracing for SARS-CoV-2 in Switzerland. Swiss Med Wkly 2020 Dec 14;150:w20457. [doi: 10.4414/smw.2020.20457] [Medline: 33327003]
27. Impact du confinement sur les français lié à la crise sanitaire Covid-19. France Soir, MIS Group. 2021. URL: http://www.francesoir.fr/lifestyle-bien-etre/accepteriez-vous-detre-trace-electroniquement-pour-eviter-la-propagation-des [accessed 2021-02-26]
28. Morley J, Cowls J, Taddeo M, Floridi L. Ethical guidelines for COVID-19 tracing apps. Nature 2020 Jun;582(7810):29-31. [doi: 10.1038/d41586-020-01578-0] [Medline: 32467596]
29. Dar AB, Lone AH, Zahoor S, Khan AA, Naaz R. Applicability of mobile contact tracing in fighting pandemic (COVID-19): issues, challenges and solutions. Comput Sci Rev 2020 Nov;38:100307 [FREE Full text] [doi: 10.1016/j.cosrev.2020.100307] [Medline: 32989380]
30. Kokolakis S. Privacy attitudes and privacy behaviour: a review of current research on the privacy paradox phenomenon. Comput Secur 2017 Jan;64:122-134. [doi: 10.1016/j.cose.2015.07.002]
31. i-Share. URL: https://www.i-share.fr/ [accessed 2021-02-26]
32. Montagni I, Qchiqach S, Pereira E, Tully PJ, Tzourio C. Sex-specific associations between sleep and mental health in university students: a large cross-sectional study. J Am Coll Health 2020 Apr;68(3):278-285. [doi: 10.1080/07448481.2018.1546183] [Medline: 30615574]
33. Pazzaglia A, Stafford E, Rodriguez S. Survey methods for educators: selecting samples and administering surveys (part 2 of 3). Appl Res Methods 2016:1-29.
34. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
35. Rogers RW. A protection motivation theory of fear appeals and attitude change. J Psychol 1975 Sep;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]
36. EpiData Entry 3.1. URL: https://www.epidata.dk/index.htm [accessed 2021-03-02]
37. Jordan RE, Adab P, Cheng KK. Covid-19: risk factors for severe disease and death. BMJ 2020 Mar 26;368:m1198. [doi: 10.1136/bmj.m1198] [Medline: 32217618]
38. Park S, Choi GJ, Ko H. Information technology-based tracing strategy in response to COVID-19 in South Korea-privacy controversies. JAMA 2020 Jun 02;323(21):2129-2130. [doi: 10.1001/jama.2020.6602] [Medline: 32324202]

Abbreviations

i-Share: Internet-Based Students Health Research Enterprise
KABP: knowledge, attitudes, beliefs, and practices
PTM: protection motivation theory
TAM: technology acceptance model

Edited by R Kukafka; submitted 10.12.20; peer-reviewed by F Rowe, T Loda; comments to author 24.12.20; revised version received 31.12.20; accepted 31.01.21; published 03.03.21.

Please cite as:
Montagni I, Roussel N, Thiébaut R, Tzourio C
Health Care Students' Knowledge of and Attitudes, Beliefs, and Practices Toward the French COVID-19 App: Cross-sectional Questionnaire Study
J Med Internet Res 2021;23(3):e26399
URL: https://www.jmir.org/2021/3/e26399
doi:10.2196/26399
PMID:33566793

Original Paper

Investigating Associations Between Screen Time and Symptomatology in Individuals With Serious Mental Illness: Longitudinal Observational Study

Philip Henson[1], MS; Elena Rodriguez-Villa[2], BA; John Torous[2], MD

[1]Mt Sinai School of Medicine, New York, NY, United States

[2]Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States

Corresponding Author:
John Torous, MD
Beth Israel Deaconess Medical Center
Harvard Medical School
75 Fenwood Rd
Office 618
Boston, MA, 02115
United States
Phone: 1 6177541213
Email: jtorous@bidmc.harvard.edu

Abstract

Background: Increasing screen time exposure from digital devices like smartphones has shown a variety of mixed associations with cognition, behavior, and well-being in adults and children but little is known about its associations with symptomatology in individuals with serious mental illness.

Objective: To determine the range of associations between screen time and symptoms of individuals with mental illness, we utilized a method called specification curve analysis.

Methods: In this observational study, we recruited smartphone-owning adults (≥18 years old) with schizophrenia and healthy controls. We installed 2 research-source smartphone apps, mindLAMP and Beiwe, to collect survey results, cognitive test results, and screen time metrics over a period of 3 months. Surveys were scheduled for twice a week, but participants were instructed to take the surveys naturally as much or as little as they wanted. Screen time was collected continuously in the background. A total of 140 participants was recruited from the outpatient clinic population as well as through general public advertising. Age-matched, smartphone-owning healthy controls were also part of the recruitment pool. A specification curve analysis was a priori designed to explore the relationship between every combination of independent variable and dependent variable in order to demonstrate to what degree screen time relates to symptoms in individuals with serious mental illness.

Results: The sample consisted of 88 participants (54 with schizophrenia and 34 healthy controls) who completed both the initial and follow-up visits, completed at least one self-reported survey, and had a minimum passive data cutoff of 5 consecutive days. While we found an association between smartphone screen time metrics and cognition (adjusted R^2=0.107, P<.001), specification curve analysis revealed a wide range of heterogenous associations with screen time from very negative to very positive. The effects differed based on diagnostic group, age bracket, type of regression model used, and the specific independent and dependent variables selected for analysis.

Conclusions: The associations between screen time and mental health in patients with schizophrenia are heterogenous when examined with methods that reduce analytical bias. The heterogeneity in associations suggests that complex and personalized potential effects must be understood in the greater context of an individual. This analysis of longitudinally collected screen time data shows potential for future research that could benefit from high resolution metrics on smartphone use.

(J Med Internet Res 2021;23(3):e23144) doi:10.2196/23144

KEYWORDS

mHealth; schizophrenia; apps; mobile; screen time

Introduction

Technology pervades many, if not most, facets of daily life. Advances in functionality, speed, customization, and smart programming offer opportunities to access, communicate, and share resources with unprecedented efficiency. But this same connectivity has also raised issues around mental health impact. Digital devices are a mode for connectivity, explaining why 74% of Americans use a computer for their work [1] and 81% of Americans own a smartphone [2]. Screen time is thus a biproduct of productivity and sociability for many people. Resulting fears around increasing time spent looking at a screen and mental health concerns—whether it is for work or for leisure—have emerged.

Concerns, for youth especially, focus on how screen time may hinder physical activity, attention, cognition, and emotional well-being. Despite the vast increases in screen time, often across multiple devices and above recommended limits [3], data to validate or obviate concerns are limited. Studies to date show contradictory evidence on short-term and long-term effects of screen time on youth and adults [4]. Small sample sizes, self-report measures, and a deficit of longitudinal research have yielded inconclusive results. A recent paper used specification curve analysis (SCA) to highlight the degree to which study results measuring the impact of screen time on youth varied based on analytical choices [5]. The analysis enumerated the associations between screen time and emotional well-being among young people but also illustrated that the effect is smaller than other influences such as bullying.

Studies involving screen time among individuals with severe mental illness are even more scarce. On one hand, screen time may be of benefit for patients with serious mental illness as it may facilitate beneficial social connections that may be lacking offline and offer patients a sense of community and connection [6]. On the other hand, screen time may distract patients with serious mental illness from real-world obligations or expose them to harmful and stigmatizing content. One study in individuals with schizophrenia showed a positive correlation between phone use and functional capacity and cognition [7], but overall research on the topic is lacking. The need for understanding on this subject is critical, as research from a 2016 study suggests that nearly 50% of people with serious mental illness may spend up to 3 hours per day in front of their phone screen and nearly 20% may spend up to 10 hours per day on a computer screen [8], potentially making screen time a key exposure in their routine and daily life.

Screen time and exposure have likely only increased for all people, including those with serious mental illness, in the last 4 years since that 2016 study. Understanding how screen time has a positive or negative effect—or any effect at all—on the well-being of individuals with serious mental illness like schizophrenia is thus critical for ensuring today's care remains responsive to the exposure and realities faced by patients. The data needed to identify problematic screen time involve total screen time, session time, and number of checks from smartphones. A recent study found that typical smartphone usage is relatively consistent and can be inferred with just 5 days of data [9]. On the other hand, habitual checking behaviors (sessions lasting less than 15 seconds) that may be indicative of preoccupation with mobile phones, can be inferred with just 2 days of data. These results, along with the finding that self-reported smartphone usage did not correlate with the objective measurements, suggest an important opportunity to use smartphone-derived screen time metrics in studying its effects.

Our research aimed to address a gap in the literature and understanding by investigating the effect of screen time on individuals with schizophrenia. Our unique dataset, with longitudinal objective screen time measures, self-report surveys, and both baseline and longitudinal cognition tests, offers an opportunity to begin to appreciate the impact of screen time on a subset of patients with serious mental illness. In relying on objective metrics of screen time derived from longitudinal phone screen on/off sensor data instead of single time-point, self-reported screen time, we hoped to avoid biases that have made prior works on screen time and mental health difficult to generalize. In this paper, we aimed to (1) investigate the association between screen time and baseline cognition in individuals with schizophrenia, (2) determine the impact of screen time on symptomatology in both people with schizophrenia and healthy controls via SCA to determine if effects hold across all possible analytical combinations, and (3) identify the association between screen time and symptomatology on an individual basis in both people with schizophrenia and healthy controls. We sought to determine the extent of screen time's effect on symptoms in individuals with schizophrenia at both group and individual levels and expected to find complicated and heterogeneous associations between screen time metrics and symptomatology.

Methods

Longitudinal Data Collection Platforms

Two types of longitudinal data were collected: (1) active data in the form of participant self-reported surveys and cognitive tests and (2) passive data that included GPS, accelerometer, and screen time. Two research applications, mindLAMP and Beiwe, were installed on participants' smartphones after receiving institutional review board approval at the Beth Israel Deaconess Medical Center [10,11].

Participants

For both studies, smartphone-owning adults (≥18 years old) were recruited from the greater Boston area starting August 2018 through the Massachusetts Mental Health Center in Boston, MA and general public advertising for convenience sampling of controls. A total of 140 participants enrolled after signing written informed consent, 6 dropped out, and 46 were excluded for not providing at least one self-reported survey or having inadequate screen time data (a minimum of 5 days of smartphone usage was used as a passive data cutoff). Of the 88 remaining participants, 54 had a clinical diagnosis of schizophrenia (SZ), and 34 were healthy controls (HC). All participants owned a smartphone and were given a smartwatch for the duration of the study to assist in data collection. Demographic information can be found in Table 1.

Table 1. Demographic characteristics of 88 smartphone-owning participants from the greater Boston area.

Characteristics	HC[a] (n=34)	SZ[b] (n=54)	*P* value
Age (years), mean (SD)	39.62 (14.56)	33.02 (11.71)	.250
Gender, n (%)			.681
Male	19 (56)	25 (46)	
Female	13 (38)	25 (46)	
Other	2 (6)	4 (7)	
Race, n (%)			<.001
American Indian or Alaskan Native	0 (0)	4 (7)	
Asian	27 (79)	3 (6)	
Black or African American	2 (6)	19 (35)	
Multiracial or other	1 (3)	1 (2)	
Native Hawaiian or Pacific Islander	0 (0)	1 (2)	
White Caucasian	4 (12)	20 (37)	
Not reported	0 (0)	6 (11)	
Education, n (%)			<.001
Some high school	0 (0)	2 (4)	
High school graduate or GED[c]	3 (9)	15 (28)	
Some college	2 (6)	20 (37)	
4-year college graduate or higher	29 (85)	17 (32)	

[a]HC: healthy control.

[b]SZ: clinical diagnosis of schizophrenia.

[c]GED: General Educational Development test.

Data Collection Protocol

After signing informed consent, participants completed paper-and-pencil symptom surveys, completed a cognitive assessment with a validated iPad version of the Brief Assessment of Cognition in Schizophrenia (BACS, SZ group only) [12], and installed mindLAMP and Beiwe on their smartphones. BACS was not administered for the HC group due to the lack of a psychiatric diagnosis as well as the assessment's specificity for individuals with schizophrenia. For 3 months, participants were notified on their smartphones to take 10 surveys per week: 2 each of mood (PHQ-9 [Patient Health Questionnaire-9]) [13], anxiety (GAD-7 [7-item Generalized Anxiety Disorder assessment]) [14], sleep, and sociability. Each survey ended with a cognitive test: Jewels A or Jewels B, which are smartphone-adapted versions of the classic Trails-A and Trails-B tasks to assess a wide variety of cognitive domains including attention, visual search, task switching, and psychomotor speed [15]. Jewels B was used for the analysis as it is a more complex task and has shown better separation between individuals with psychosis and healthy controls [16]. A single score, or "beta value," was used to represent performance on a Jewels task and takes into account both accuracy and error rate. Meanwhile, the Beiwe app collected multiple passive data streams (GPS, accelerometer, screen on/off, and call and text logs) simultaneously and uploaded the data to a Health Insurance Portability and Accountability Act–compliant server every hour. Raw screen time (in seconds) was calculated by summing the intervals between "Screen On" and "Screen Off" data points. While participants were paid for their clinical visits and in-person surveys, no study compensation was provided for app engagement or survey completion.

Data Analysis

All analyses were performed using the R programming language (version 3.6.2 [17]). Raw screen time data were aggregated by day and processed into 3 main screen time metrics: (1) screen time (seconds), (2) session time (seconds), and (3) number of checks (unitless). Session time was calculated by dividing screen time by the number of sessions and checks (ie, "habitual checking behaviors" were sessions lasting less than 15 seconds) [9]. Smartphone surveys were also aggregated by day, and survey scores were averaged if more than one of the same survey was taken that day (eg, separate PHQ-9 results of 10 and 11 on the same day would be converted to 10.5).

Correlations between the first month of screen time and baseline cognition were conducted using the Spearman rank correlation coefficient, and *P* values were adjusted using the false discovery rate method. Multivariate multiple linear regression was performed on longitudinal screen time and Jewels B cognition beta values.

SCA was inspired by Orben and Przybylski [5] and aided by the "specr" package in R [18]. Gender (male or female) was added as a covariate, and 2 models were used: linear model and generalized linear model. Groups were separated for SCA based

on age (<30 years old or ≥30 years old) and diagnostic group (SZ or HC).

Linear model regression was performed at the individual level between screen time and symptoms, and regression estimates were arranged from high to low. Only individuals with survey results in each of the 4 survey categories of interest (mood, anxiety, sleep, social) were included in this portion of the analysis.

Results

Cognition

Participants with schizophrenia (n=54) were assessed at baseline for cognition via BACS, and all 88 participants (54 SZ and 34 HC) were assessed longitudinally via the Jewels B assessment within the mindLAMP smartphone app.

Baseline

Among the 6 subdomains of the BACS (Verbal Memory, Verbal Fluency, Digit Sequencing, Symbol Coding, Token Motor, and Tower of London), Spearman correlations between screen time metrics and SZ baseline BACS subdomains ranged from –0.17 to 0.29, but there were no significant correlations (*P* values ranged from .25 to .98).

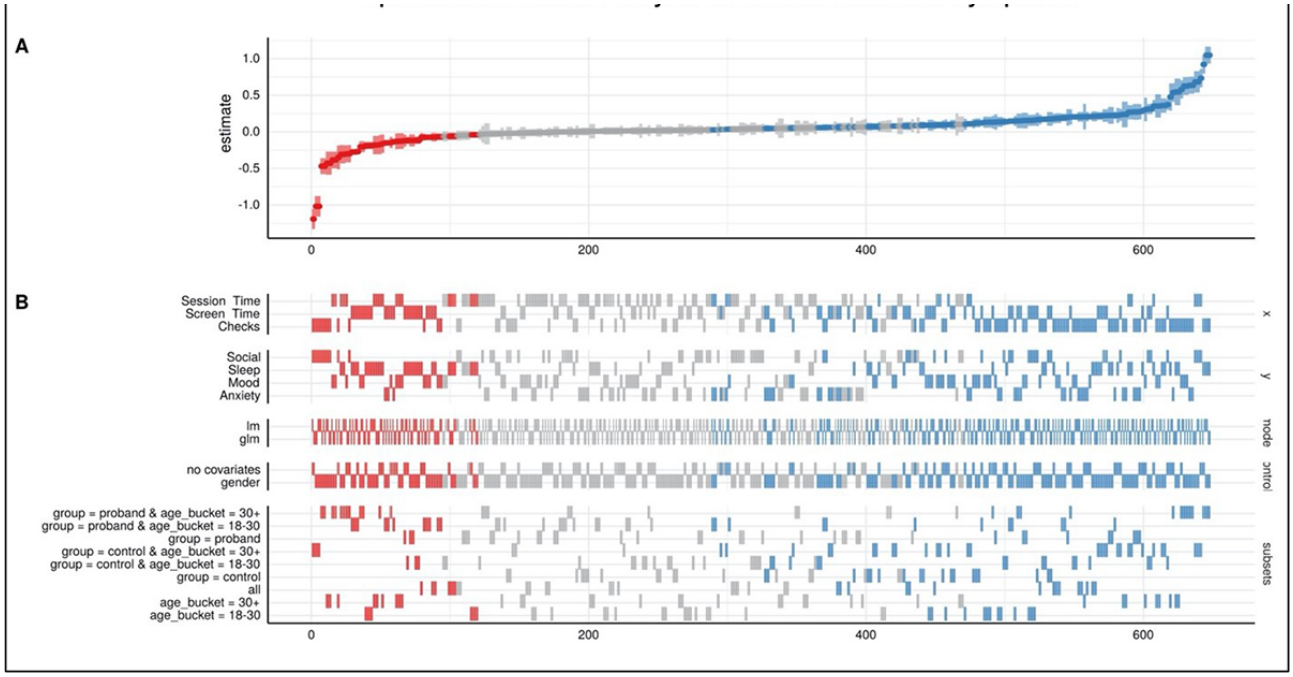
Longitudinal

Multivariate multiple linear regression revealed a significant regression equation in SZ for the association of the screen time metrics (number of sessions, number of checks, screen time, and session time) with longitudinal beta values for cognition (Jewels B assessment [16]): $F_{4,144}$=5.43, *P*<.001 with an adjusted R^2 of 0.107. The beta values used to represent cognition take into account accuracy and error rate, and the greater the beta value, the better the performance. The regression equation for HC was not significant (adjusted R^2=0.035, *P*=.068).

Specification Curve Analysis

SCA on data for all 88 participants (54 SZ and 34 HC) revealed estimates (regression coefficient β) for over 600 combinations, or specifications, ranging from –1.19 to 1.05 (Figure 1). The figure is meant to be a high-level representation of the heterogeneity of associations between screen time metrics and symptoms, displayed in order of most negative on the left to most positive on the right. If we were to zoom in, we could see, for example, that an individual column within the red area might involve screen time, sociability, linear model, covarying for gender, and including participants with schizophrenia over the age of 30 years. For that group, the model found a significant negative association between self-reported sociability behavior and smartphone screen time. Individual analyses can be read as a vertical column, with each column (ie, specification) representing a unique combination of variables that was tested in this analysis. For example, the leftmost column of results involves checks, social, linear model, no covariates, and the control group of participants aged ≥30 years. In other words, this is the most negative association (β=–1.19) and suggests that more checking behavior was associated with worse reported sociability for an older control group while not adjusting for gender using a linear model. On the right side of the plot, the most positive association (β=1.05) was between checks and sleep (ie, more checking behavior was associated with better reported sleep) in SZ, but not HC, for individuals over 30 years old using a generalized linear model and adjusting for gender.

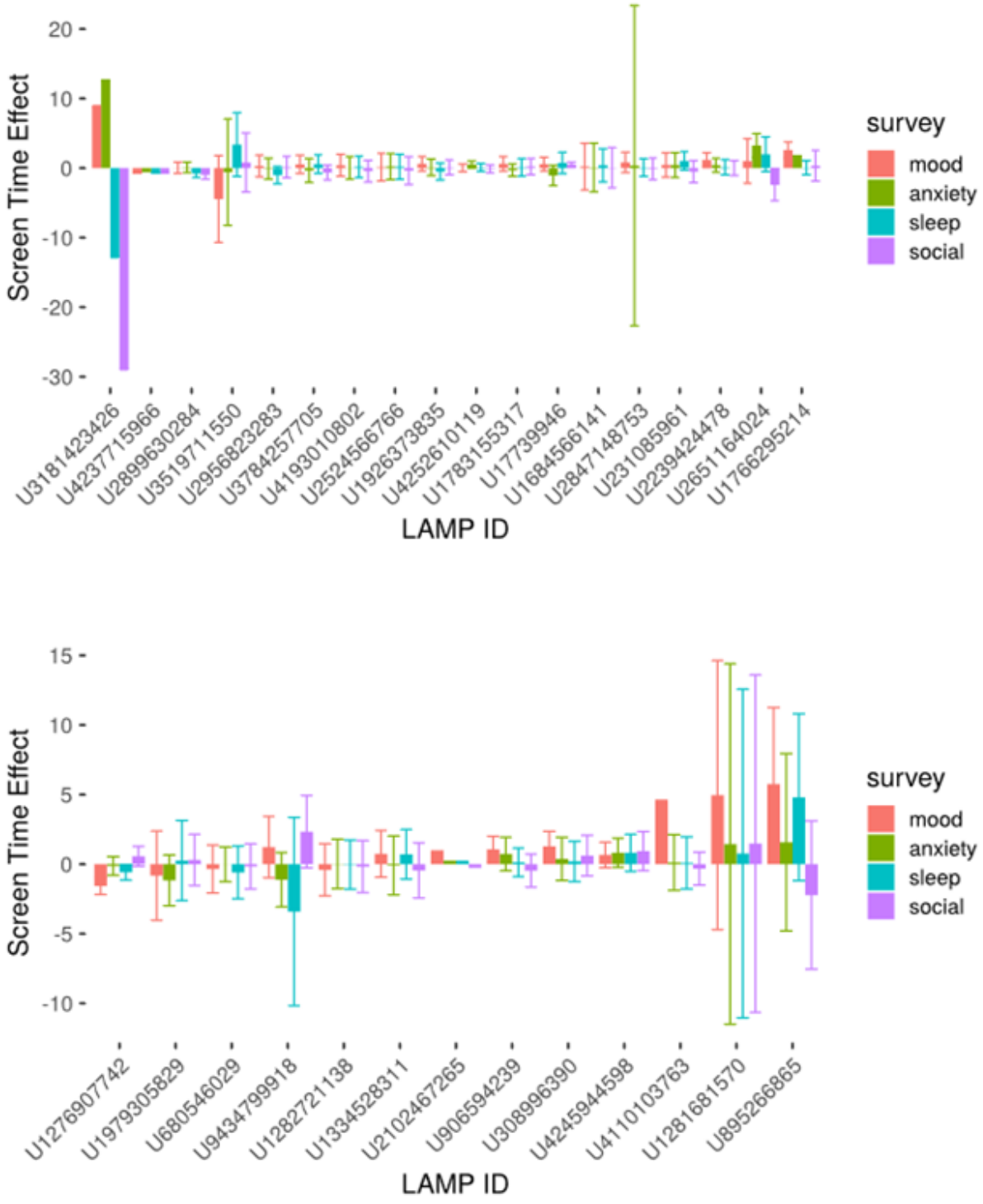
Figure 1. Specification curve analysis of screen time and symptoms, as a visual representation of the >600 combinations of regression analyses between screen time metrics and self-reported mood, anxiety, sleep, and sociability: (A) regression coefficient, (B) variables.



Individual Participant Regression

As shown in Figure 2, linear model estimates at the participant level revealed differences in screen time's effect on symptoms among individuals as well as between groups, with lower and more negative values identified in the SZ group (mean –0.136, SD 4.3) than in the HC group (mean 0.51, SD 1.6). Individual estimates ranged in SZ from –28.9 for sociability to 12.6 for anxiety (mean 0.51, SD 1.6) and in HC from –3.4 to 5.8 in HC (mean –0.136, SD 4.3). Note that only participants with results in all 4 survey domains were included in this analysis (18 SZ and 13 HC), and completion of surveys was optional in the study.

Figure 2. Individual effects of screen time on symptoms measured using the mindLAMP mobile phone app, plotted from low to high for (A) participants with a clinical diagnosis of schizophrenia (n=18) and (B) healthy controls (n=13), with self-reported survey scores in all categories.



Discussion

Principal Findings

Using longitudinal and objective measurements related to screen time, we found a range of positive and negative associations with mental health symptoms at both population and individual levels in patients with schizophrenia. We found no significant associations between screen time and baseline cognition but did find associations in schizophrenia between cognition as measured by the Jewels B cognitive task and screen time metrics (adjusted R^2=0.107, P<.001). The low R^2 indicates that, despite the significant trend, much of the variability cannot be explained by the model. Nonetheless, patterns of smartphone use may be related to underlying cognitive functioning—presenting opportunities for capturing novel data and designing more engaging apps.

However, it is important to consider the multitude of factors in addition to screen time that may contribute to well-being, symptomatology, and cognition. An SCA on 3 large data sets of over 350,000 people found that many other personal and behavioral factors can have as much, if not more, of an effect on well-being than technology use [5]. For example, bullying and marijuana use had a greater negative effect on well-being than technology use, whereas proper sleep and nutrition had a greater positive effect. The authors even drew attention to neutral factors like eating potatoes as having nearly as negative of an effect on technology use. Our dataset is much smaller but does rely on objective metrics of screen time and also suggests that at a population level, screen time itself is not highly associated with symptomology.

As with previous screen time analysis studies, our results show a wide range of regression results based on the chosen analysis and isolated components. However, the SCA did bring to light some interesting results around group differences in the study. For example, both screen time and checking behavior had greater positive associations with symptoms of older controls than younger controls (x2.1 and x36.9, respectively) and checking time across all ages in the SZ group had a more positive association with symptoms than in the HC group (x3.48). While there are many potential reasons for these associations beyond the scope of this study, they highlight potential beneficial associations contrary to some perceptions that older adults or those with schizophrenia may not want to engage or even use technology. Results that screen time and checking behaviors were associated with improved sleep outcomes in those with schizophrenia but not controls also highlights that results in healthy controls do not always mirror those in patients, and that caution is necessary if seeking to apply the broader research base of screen time in the general population to those with serious mental illness.

On an individual basis, our results suggest that simple rules or guidance around screen time and mental health for individuals with schizophrenia, or controls, may not be practical. The individual participant analysis revealed heterogeneity in the effect of screen time on symptoms, with individual screen time effect estimates ranging in the SZ group from –28.9 for sociability to 12.6 for anxiety (mean 0.51, SD 1.6) and from –3.4 to 5.8 in the HC group (mean –0.136, SD 4.3). In this sample, the effects on the HC group were more positive but less variable than in the SZ group. This could be because the controls in this smartphone study may be naturally more interested in technology and thus more technology-literate and use their smartphones more regularly in ways that improve symptoms, for example playing a game to reduce stress. In addition, the greater variability in the SZ group may be due to inherent variability in symptoms and behavior associated with serious mental illness that is not present in healthy controls.

There are very few studies investigating neurocognitive effects of screen time in schizophrenia, and our results do not yet suggest conclusive evidence to fill this literature gap. Our baseline assessment of cognition was not associated with any screen time metrics, but this analysis was performed at a population level, and results may be washing out due to individual variance. There is, however, evidence that past mobile phone use in older (>40 years old) individuals with schizophrenia is associated with higher global cognitive performance [7]. Despite the mean age difference, this is in line with our longitudinal assessment finding of an association between phone checking behavior and higher performance on the Jewels B cognitive task.

While several individual studies have found associations between screen time and symptoms of depression or anxiety in both adults and children [19-21], our SCA results are in agreement with a recent large dataset analysis (n=355,358) suggesting a complex relationship between screen time and symptoms [5]. We did not have data on other personal and behavior factors (eg, bullying, nutrition) for comparison, but the range of estimates produced by the SCA analysis demonstrates that the effect of screen time is still heterogenous and difficult to parse given all the variables.

Limitations

As with all studies, there are several limitations that need to be addressed. First, with longitudinal behavioral data, it is important to note that behavior and symptomatology can change throughout the day, so while daily aggregation employed in our study represents a high resolution for longitudinal analysis, it could be higher to capture fluctuations within the day. Still, compared to survey studies where patients estimate their mental health and screen time over weeks or even months, our methods offer improvement. Second, individual analysis only included participants with data in all survey categories. We are likely missing information from those less engaged with the app or who elected to take only a subset of surveys due to the naturalistic study design. Third, there may be a time lag between screen time and symptom change that is not accounted for as independent and dependent variables are paired for a given day, although such an effect has not yet been well characterized in the research to date. Finally, while controls were matched on age, they differed in education and race, potentially due to convenience sampling, which may have had a confounding effect.

Conclusion

Increasing screen time is a growing concern, and despite recent research efforts, there are very few studies reporting the effect of screen time on individuals with schizophrenia. The importance of a priori analysis and transparent methods around digital mental health is also highlighted in our results, which show how divergent conclusions can be supported if using more limited analysis. Our results show that variance at the individual and population levels can account for drastically different reporting of screen time's associations with symptoms, from very negative to very positive, demonstrating a complex relationship that requires further exploration.

Conflicts of Interest

None declared.

References

1. Mamedova S, Pawlowski E. A Description of U.S. Adults Who Are Not Digitally Literate. U.S. Department of Education. 2018 May. URL: https://nces.ed.gov/pubs2018/2018161.pdf [accessed 2021-02-11]
2. Mobile FS. Mobile Fact Sheet. Pew Research Center. 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-02-11]
3. Fakhouri THI, Hughes JP, Brody DJ, Kit BK, Ogden CL. Physical activity and screen-time viewing among elementary school-aged children in the United States from 2009 to 2010. JAMA Pediatr 2013 Mar 01;167(3):223-229. [doi: 10.1001/2013.jamapediatrics.122] [Medline: 23303439]
4. Stiglic N, Viner RM. Effects of screentime on the health and well-being of children and adolescents: a systematic review of reviews. BMJ Open 2019 Jan 03;9(1):e023191 [FREE Full text] [doi: 10.1136/bmjopen-2018-023191] [Medline: 30606703]
5. Orben A, Przybylski AK. The association between adolescent well-being and digital technology use. Nat Hum Behav 2019 Feb 14;3(2):173-182. [doi: 10.1038/s41562-018-0506-1] [Medline: 30944443]
6. Naslund JA, Bondre A, Torous J, Aschbrenner KA. Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice. J Technol Behav Sci 2020 Sep 20;5(3):245-257. [doi: 10.1007/s41347-020-00134-x] [Medline: 33415185]
7. Depp CA, Harmell AL, Vahia IV, Mausbach BT. Neurocognitive and functional correlates of mobile phone use in middle-aged and older patients with schizophrenia. Aging Ment Health 2016 Mar 13;20(1):29-35 [FREE Full text] [doi: 10.1080/13607863.2015.1008987] [Medline: 25768842]
8. Gay K, Torous J, Joseph A, Pandya A, Duckworth K. Digital Technology Use Among Individuals with Schizophrenia: Results of an Online Survey. JMIR Ment Health 2016 May 04;3(2):e15 [FREE Full text] [doi: 10.2196/mental.5379] [Medline: 27146094]
9. Wilcockson TDW, Ellis DA, Shaw H. Determining Typical Smartphone Usage: What Data Do We Need? Cyberpsychol Behav Soc Netw 2018 Jun;21(6):395-398. [doi: 10.1089/cyber.2017.0652] [Medline: 29781716]
10. Torous J, Wisniewski H, Bird B, Carpenter E, David G, Elejalde E, et al. Creating a Digital Health Smartphone App and Digital Phenotyping Platform for Mental Health and Diverse Healthcare Needs: an Interdisciplinary and Collaborative Approach. J. technol. behav. sci 2019 Apr 27;4(2):73-85 [FREE Full text] [doi: 10.1007/s41347-019-00095-w]
11. Torous J, Kiang MV, Lorme J, Onnela J. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. JMIR Ment Health 2016 May 05;3(2):e16 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
12. Atkins AS, Tseng T, Vaughan A, Twamley EW, Harvey P, Patterson T, et al. Validation of the tablet-administered Brief Assessment of Cognition (BAC App). Schizophr Res 2017 Mar;181:100-106 [FREE Full text] [doi: 10.1016/j.schres.2016.10.010] [Medline: 27771201]
13. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
14. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
15. Salthouse TA. What cognitive abilities are involved in trail-making performance? Intelligence 2011 Jul;39(4):222-232 [FREE Full text] [doi: 10.1016/j.intell.2011.03.001] [Medline: 21789028]
16. Liu G, Henson P, Keshavan M, Pekka-Onnela J, Torous J. Assessing the potential of longitudinal smartphone based cognitive assessment in schizophrenia: A naturalistic pilot study. Schizophr Res Cogn 2019 Sep;17:100144 [FREE Full text] [doi: 10.1016/j.scog.2019.100144] [Medline: 31024801]
17. The R Project for Statistical Computing. The R Foundation. 2021. URL: https://www.r-project.org/ [accessed 2021-02-12]
18. Masur P, Scharkow M. specr: Statistical functions for conducting specification curve analyses. GitHub. 2020. URL: https://github.com/masurp/specr [accessed 2021-02-11]

19. Madhav K, Sherchand SP, Sherchan S. Association between screen time and depression among US adults. Prev Med Rep 2017 Dec;8:67-71 [FREE Full text] [doi: 10.1016/j.pmedr.2017.08.005] [Medline: 28879072]
20. Maras D, Flament MF, Murray M, Buchholz A, Henderson KA, Obeid N, et al. Screen time is associated with depression and anxiety in Canadian youth. Prev Med 2015 Apr;73:133-138. [doi: 10.1016/j.ypmed.2015.01.029] [Medline: 25657166]
21. Kremer P, Elshaug C, Leslie E, Toumbourou JW, Patton GC, Williams J. Physical activity, leisure-time screen use and depression among children and young adolescents. J Sci Med Sport 2014 Mar;17(2):183-187. [doi: 10.1016/j.jsams.2013.03.012] [Medline: 23648221]

Abbreviations

BACS: Brief Assessment of Cognition in Schizophrenia
GAD-7: 7-item Generalized Anxiety Disorder assessment
HC: healthy control
PHQ-9: Patient Health Questionnaire-9
SCA: specification curve analysis
SZ: clinical diagnosis of schizophrenia

Edited by R Kukafka; submitted 02.08.20; peer-reviewed by S Bradstreet, L Balcombe; comments to author 25.08.20; revised version received 01.09.20; accepted 31.01.21; published 10.03.21.

Please cite as:
Henson P, Rodriguez-Villa E, Torous J
Investigating Associations Between Screen Time and Symptomatology in Individuals With Serious Mental Illness: Longitudinal Observational Study
J Med Internet Res 2021;23(3):e23144
URL: https://www.jmir.org/2021/3/e23144
doi:10.2196/23144
PMID:33688835

Original Paper

Development of a Smartphone App to Predict and Improve the Rates of Suicidal Ideation Among Transgender Persons (TransLife): Qualitative Study

Alex Dubov[1*], PhD; Liana Fraenkel[2*], MPH, MD; Zil Goldstein[3*], FNP-BC, AAHIVS; Hansel Arroyo[4*], MD; Derek McKellar[5*], MBA; Steve Shoptaw[6*], PhD

[1]School of Behavioral Health, Loma Linda University, Loma Linda, CA, United States

[2]School of Medicine, Yale University, New Haven, CT, United States

[3]Department of Medicine, Callen-Lorde Community Health Center, New York, NY, United States

[4]Icahn School of Medicine, Mount Sinai Hospital, New York, NY, United States

[5]Accenture LLC, Los Angeles, CA, United States

[6]Department of Family Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[*]all authors contributed equally

Corresponding Author:
Alex Dubov, PhD
School of Behavioral Health
Loma Linda University
11065 Campus Street
Loma Linda, CA, 92350
United States
Phone: 1 9095581900
Email: adubov@llu.edu

Abstract

Background: Transgender people are at a high risk of suicidal ideation, suicide attempts, and deaths. Among transgender individuals, 77% and 41% engage in suicidal ideation and suicide attempt in their lifetime, respectively, which exceeds the general population rates (9.2% and 2.7%, respectively). Traditionally, suicide risk factors have been studied over a long period between measurements, making it difficult to understand the short-term variability in suicide risk. Mobile phone apps offer an opportunity to understand the immediate precursors of suicidality through the assessment of behaviors and moods in real time. This is the first study to use a mobile phone app (TransLife) to understand the short-term risk factors for suicide among transgender individuals.

Objective: This study aims to beta test the usability of an evidence-informed mobile health (mHealth) suicide prevention phone app, TransLife. The primary aims are to obtain preliminary data on user engagement and satisfaction with the app, and to assess the feasibility of completing ecological momentary assessments (mood logs) within the app.

Methods: We used qualitative methods and an exploratory research approach that combined naturalistic app use, focus groups, and semistructured phone interviews. The focus group was informed about the development of the prototype. We conducted a 3-week evaluation to determine engagement and obtain detailed user feedback about the app. After participation in the pilot, phone-based, semistructured, and audio-recorded exit interviews were conducted with the research participants.

Results: In total, 16 transgender individuals participated in this study. On average, users logged in 4 (SD 2.7) times a week and spent approximately 5 (SD 3.5) minutes on the app per log-in. A total of 6 major themes emerged in this study. These themes focused on the app's functionality, satisfaction with using the app, perceived ease of use, perceived safety of providing personal data within the app, trusting the app enough to share personal feelings, and features that make this app engaging. These themes suggest that TransLife is an engaging, useful, and acceptable mHealth intervention. Participants reported that the app was easy to use and understand, supported mental self-care, promoted self-awareness, and helped them identify triggers of negative moods.

Conclusions: The results of this pilot study indicate that TransLife is an engaging, acceptable, and potentially effective mHealth intervention. Transgender participants reported many advantages of using TransLife, such as being able to track their mood, connecting to the community, and accessing local resources. This study provides initial support for the acceptability and usability of TransLife as an mHealth intervention designed for the transgender community.

(J Med Internet Res 2021;23(3):e24023) doi:10.2196/24023

KEYWORDS

mobile health; mHealth; mobile app; pilot study; qualitative research; user-centered design; acceptability study; health services for transgender persons; suicide prevention; mental health; mobile phone

Introduction

Background

Transgender people, who represent up to 0.5% of the adult population, have an extremely high prevalence of suicidal ideation (SI; 77%) [1] and suicide attempts (SAs; 41%) [2]. The rate of lifetime SI and SAs exceeds that of the overall US population (9.2% and 2.7%, respectively) and the lesbian, gay, and bisexual population (10%-20%) [3]. The majority of previous research on SI and SAs among transgender people has focused on the same sociodemographic (sex assigned at birth, employment, age, and ethnicity) and psychiatric risk factors (eg, depression) for suicide. However, these factors are not specific to SI and SAs, do not indicate when an individual is at increased risk, and are removed in time from an actual SA. Furthermore, existing studies of SI and SAs among the transgender population have a number of methodological limitations in how studies define transgender people (eg, exclusion of nonbinary) or measure SI and SAs (eg, a mixture of validated and researcher-designed measures) and in their mostly cross-sectional nature. Current models of suicidal behavior among transgender people have a limited ability to predict suicide risk. Improved methods to identify transgender persons at a high risk of SI are urgently needed.

Recent research has demonstrated that SI fluctuates over time and varies considerably within each person [4]. From a prevention perspective, it is crucial to understand the factors that are the strongest predictors of SI in the weeks, days, and hours before an SA. Long-term risk factors of SAs, described in the literature, are poor predictors of short-term risk [5]. This is one of the most significant limitations of the existing studies, namely, their tendency to measure SI risk at a single point in time, covering long retrospective periods. Currently, no studies have proposed an understanding of short-term modifiable risk factors among the transgender population. To address this limitation, it is necessary to capture new, dynamic, and real-time data on immediate risk factors for SI. Mobile technology allows researchers to augment standard surveys and clinical examinations with dynamic, real-time data. Ecological momentary assessment (EMA) embedded in phone apps offers the opportunity to understand the immediate precursors of SI in daily life by assessing behaviors, experiences, and mood states in real time.

Objectives

This study aims to beta test the usability of an evidence-informed mobile health (mHealth) SI prevention phone app, TransLife. The TransLife app provides transgender people with resources to promote resilience and cope with minority stressors from the Meyer Minority Stress framework [6]. The framework is based on the premise that the experiences of stigmatization (eg, prejudice, concealment, self-stigma, expectation of rejection) take the form of a specific minority stressor, which in turn affects the state of health of a transgender person [7]. The TransLife app also enables longitudinal data collection that can be used to create a predictive model of SI risk among transgender people. Before we finalized the development of the TransLife, we conducted a focus group with several transgender community leaders to ensure usability. This paper describes the development of the TransLife app and its features and the results of beta testing research. Beta testing aims to test software with end users in a real-world setting to identify and rectify any potential issues before being released. To ensure that a mobile app is useful, experts recommend that it should be evaluated in terms of its utility (whether an app provides the features the users need) and usability (how easy and pleasant the features are) with users. In this study, we aim to obtain an in-depth understanding of users' acceptance of the TransLife app to monitor moods and feelings, track health activities, and connect with the broader community.

Methods

App Development

A multidisciplinary team of software engineers, transgender health professionals, mHealth specialists, and psychiatry specialists created an initial version of the TransLife app. This team carried out an iterative [8], user-centered [9] design process to develop the TransLife app prototype with periodic input on concepts, features, and measures from the Transgender Health Program team of St. John's Well Child and Family Center in Los Angeles. All St. John's team members identify as transgender and represent target end users, in addition to bringing a wealth of experience in providing health and social services to more than 1400 transgender patients. Once a prototype app (version 1.0) was completed, it was further tested in a focus group discussion with staff and clients of St. John's Transgender Health Program to identify preferences and requirements to consider including in the TransLife. Their feedback was incorporated into the design of version 2.0 beta tested for usability and acceptability in this study.

The App

The TransLife app has several components: *Dashboard* is the landing page of the app. It provides users with updates and information regarding their progress, surveys that need to be completed, upcoming events, and reminders. It offers important notifications and announcements to help users use apps more effectively. It also dynamically suggests user resources and activities based on their mood patterns collected through the app's usage. For instance, if the mood has been persistently negative for the past 3 or more days, the app will recommend taking a survey (brief measures of depression and SI) typically administered within 2-week intervals. Users will also receive practical tips on coping with negative moods and positive

reinforcement when their mood improves. *Mood logger* allows users to log their mood, providing data regarding the events in a user's life and the corresponding effect on their mood. Users can access their long-term mood tracking data within the app, whereas researches collect the same well-structured data on the backend. Users can record their general mood, several related emotions, and potential reasons for this emotional state (together with their notes). This feature design is intuitive, colorful, and easy to use. Users are invited to log their sleep quality and energy levels. Users and researchers can create and publish *events*; they can search through the list of useful resources (health providers, services, etc) vetted by other app users in the user's vicinity on the *local* page (similar to the Yelp feature). The *Insights* tab displays a log history of the user's mood. Logs can be seen by week, month, and year to track progress. This feature provides a graphical representation of moods to quickly assess progress. Minimal demographic data were collected during the sign-up process to streamline the first experience with the app. The *Surveys* feature allows researchers to gather more relevant demographic data in addition to periodic monitoring of users' depressive moods and SI (at minimum every 2 weeks, when a reoccurring reminder is pushed to the user). Finally, there is a list of *resources* under a dedicated tab divided into health care and mental health resources, housing and legal resources, and resources for victims of violence. This feature has a comprehensive search and filter option to help the user get to the topic they need.

Safety Features

We ensured that the app and related data could be completely wiped from a device when the participant leaves the study, and we also provided assurances to the participant. All secondary agreements (eg, run-time libraries, standard services provided by the carrier) that collect and send data to third parties were identified and evaluated for risk. Files and data stored by the app were automatically encrypted whenever the device was locked. The latest version of secure socket layer (SSL; version 3.0) / transport layer security (TLS; version 1.3) encryption protocol was used for all communications between the app and other systems, including user authentication and the transfer of sensitive information. Data were stored and backed up using encrypted, password-protected storage services provided by the university.

Safety planning is a clinical intervention component often included in the treatment of suicidal patients and specifically aimed at the transition from having the thoughts or intention to engage in an SA and acting upon the thoughts or intention [10]. The use of safety planning is recommended for patients at risk of SAs in suicide prevention guidelines [11], and safety planning is embedded in cognitive behavioral therapy (CBT) for suicide prevention [12]. Our safety plan conforms to the best practices [13] and includes the following steps: (1) identifying early warning signs of an impending suicidal crisis (eg, negative feelings and problematic behaviors), (2) employing internal coping strategies, (3) employing distraction activities and socialization to distract from SI, (4) making use of social support contacts who may offer help, (5) collating the contact details of mental health professionals and other crisis resources, and (6) making the environment safe. The Safety Plan feature is a part of onboarding new users, together with informed consent. The Safety Plan feature is stored in the app.

Study Design

This pilot study used qualitative methods and an exploratory research approach that combined naturalistic app use, focus groups, and semistructured phone interviews. The primary aims were to obtain preliminary data on user engagement and satisfaction with the TransLife app and to assess the feasibility of completing EMAs (mood logs) within the app. To achieve these aims, we conducted a 3-week evaluation to determine engagement and obtain detailed user feedback about the app. After participation in the pilot, phone-based, semistructured, and audio-recorded exit interviews were conducted with 10 transgender participants. All interviews were transcribed and independently coded by 2 study team members. Content analysis was conducted to identify data categories and overarching themes.

Recruitment

As not all transgender individuals may publicly identify as such, it is difficult to recruit a sample using traditional techniques. One approach that has shown promise in sampling hidden populations, such as transgender people, is respondent-driven sampling (RDS) [14]. RDS relies on chain recruitment but limits the number of individuals who can be recruited by each participant. Eligible participants were recruited through transgender key informants (1 transgender woman and 1 transgender man). These informants were invited to recruit 2 additional participants. The participants in the first wave were asked to invite up to 2 peers to participate. Participants were recruited between May and June 2019. Eligibility criteria included self-identifying transgender people aged ≥18 years, living in the Greater Los Angeles area, and owning a smartphone. Interested participants were asked to complete an electronic consent form and download the TransLife app from the TestFlight (a tool that allows developers to test an app before it goes on the web on the app store). Once the app was installed onto their phone, participants were asked to click on different app features to become familiar with their functions. Participants were encouraged to use the app during the 3-week trial period; however, there was no required amount of time to use the app during the trial. Participants were compensated at the end of the trial period with a US $20 virtual Amazon gift card. The Loma Linda University Institutional Review Board approved this study.

Procedures

The System Usability Scale (SUS) [15] and the Standardized User Experience Percentile Rank Questionnaire (SUPR-Q) [16] are 2 validated and widely used measures of user experience. SUS is an industry-standard scale used to evaluate a variety of products and services, including websites, mobile phones, and computer software. SUPR-Q is a well-validated 8-item instrument that measures the quality of a website's user experience. Questions from both instruments were used to inform the development of the interview guide for this study. The final interview guide consisted of 12 items (Multimedia Appendix 1). Semistructured, phone-based, and audio-recorded

interviews were conducted with 16 participants. Nielsen et al [17] showed that conducting usability testing with only 5 participants will reveal 85% of usability problems. All interviews were conducted by one investigator (AD) who had experience with qualitative research. The interviews were transcribed and supplemented with the previous focus group discussion transcripts. Data saturation was perceived through recurring ideas, such as convenience of use, motivation to use the app, engagement with the app, and ease of use.

Data Analysis

We used a 6-step thematic analysis approach [18] to capture the user experience themes. For step 1, we familiarized ourselves with the transcripts. For step 2, we confirmed the selection of codes and themes and made necessary amendments to reach a consensus. Initial coding was performed in Atlas.ti based on the deductive coding framework, with varied responses interpreted inductively into new codes as needed. Codes were matched between authors to ensure consistency and confirm the definition of the full set of themes. In step 3, we clustered nodes into a common theme based on coherent patterns. We used some of the quotes in the Results section to demonstrate the legitimacy of the defined themes. For step 4, we reduced the themes into the most prevalent implicit and explicit ideas while deleting redundant themes. In step 5, we described the names and parameters of the identified themes. For step 6, we reported the analysis performed. In total, 6 major themes comprising several categories and codes emerged. These themes focused on the app's functionality, satisfaction with using the app, perceived ease of use, perceived safety of providing personal data within the app, trusting the app enough to share personal feelings, and features that make this app engaging. We applied methodological data triangulation by comparing data sources, namely, qualitative individual interviews and focus groups. Triangulation enhanced data richness or completeness by enabling researchers to explore a broad range of perspectives and compare and contrast perspectives on the usability and feasibility of the TransLife app [19]. The multidisciplinary team involved in the analysis helped to minimize the risk of bias because of a variety of expertise and interests by challenging each other's assumptions and interpretations.

Results

Demographic

A total of 16 transgender individuals participated in this study. Participant characteristics are shown in Table 1; participants' mean age was 33.4 (SD 9.5) years. Half (n=8, 50%) of the participants were identified as male-to-female transwomen, 37% (n=6) as female-to-male transmale, and 12% (n=2) of participants were identified as third gender (other). Participants were White (n=7, 44%), Hispanic or Latina (n=5, 31%), African American (n=2, 12%), and Asian (n=2, 13%). Most participants were highly educated (n=11, 69% of college graduates), and 63% (n=10) were employed full time.

Table 1. Demographics (N=16).

Demographic	Value
Age (years), mean (SD)	33.4 (9.5)
Gender identity, n (%)	
Male to female	8 (50)
Female to male	6 (37)
Other	2 (12)
Race or ethnicity, n (%)	
White	7 (44)
Hispanic or Latina	5 (31)
African American or Black	2 (12)
Asian or Pacific Islander	2 (12)
Education, n (%)	
High school	1 (6)
Some college	4 (25)
College graduate	7 (44)
Postgraduate studies	4 (25)
Employment status, n (%)	
Employed (full time or part time)	10 (62)
Student	3 (19)
Unemployed	3 (19)

System Usage Data

On average, users logged in 4 times a week and spent approximately 5 minutes on the app per log-in. The mood logger and question of the day were the most used features of the app. It took participants, on average, one and a half minutes to log their moods.

Engagement

Although the adoption of an app is important, users' continued interaction with an app (engagement) is key to user retention. According to a recent survey, 25% of installed mobile apps are never used, and 26% of installed mobile apps are abandoned after the first use [20]. The literature defines engagement as a psychological process that leads to the formation of loyalty [21] or the intensity of an individual's participation in and connection with app features or activities [22]. Most participants reported that they found TransLife to be helpful and engaging through its features and offerings. Participants noted some improvements in their moods and growing self-awareness as the result of using TransLife, as illustrated by this quote:

> *This app helped me to get in touch with my moods. At times I would stop and think – how do I feel now? Having to identify my feelings and emotions has helped me to become a bit more self-reflective.*

Similarly, another participant commented that the persistent use of the app helped him gain a better understanding of his mental health:

> *After about ten days of use, it has become clear that my mood is usually better next day after going to the gym the previous afternoon...there also seems to be a correlation between my low moods and unhealthy food choices...I definitely feel better when I am more active. Great insights from using this app.*

Some participants mentioned that monitoring their mood has helped them regain control over their mood states and offered some measure of self-efficacy:

> *Mood logging made me more aware of my feelings, the way my mood changes over a period of time, and how to deal with mood changes. Even when I feel down, I can remind myself that these feelings are normal and will likely pass.*

Participants noted several features of the app as driving engagement and promoting self-insight. The participant's comment on using the insights feature that provides a graphic representation of mood history is as follows:

> *Insights was one of the most helpful features. It was interesting to see how your moods change over time and progressions of either positive or negative feelings.*

During the design process, we added several features with the sole purpose of promoting engagement, such as transgender-specific news or questions of the day polls. The question of the day allows users to respond to a particular question and then compare their answers with peer responses. These app features were successful in driving engagement, as illustrated by the following quote:

> *Question of the day was something that kept me engaged with the app. I really liked seeing the answers from other folks and I can only imagine how interesting it would be to check the answers once the app will add more users.*

Functionality

The functionality feature of an app refers to the user's perception of different functions within an app and how these functions are effective in reaching the user's goals [23]. Functionality usually refers to more tangible physical features, such as layout, navigation, and esthetics. Participants described the interface as intuitive and straightforward:

> *I like the general interface and how well the app is organized. I think it is very well designed. I used mood logs every day and found it very helpful.*

Although commenting on the availability of similar mood tracking apps, several participants mentioned their intention to switch to TransLife when it becomes available within the app store. Several unique functions of the app, coupled with the overall transgender specificity of the design and content, were the main reason participants were considering switching to TransLife. Participants commented on the value of a culturally tailored mental health app for the transgender community:

> *To be honest, I've been using another mood tracking for the past several months. I would consider switching to this app when it is available in the app store because of several reasons. This app is trans specific. For example, you can add "gender dysmorphia" as one of the reasons for your bad mood. Also, this app has more to offer than just a simple mood tracker – a version of yelp, resources, even trans news.*

Participants appreciated the ability to customize the app's appearance and its functionality to fit the participants' personal needs. Participants liked to customize alerts and reminders to fit individual daily routines, for example:

> *I like the idea of being able to customize alerts and reminders from the get-go. It is so useful that the app can remind you about your meds. We used to do it with my friends, reminding each other about our meds. The mood logging reminders are also customizable, and I like that.*

Privacy and Security

Mobile apps offer tremendous potential to enhance mental health, but they often lack data security and privacy mechanisms needed to increase user's and clinician's confidence in these apps. Low confidence in an app's ability to provide a private and secure environment is a significant barrier to its adoption. Concerns about the security and privacy of the information provided on the app were evident among some participants:

> *I had some difficulty answering question of the day as I don't like putting too much information about myself out there. Maybe I am old-fashioned that way.*

However, this concern about privacy and security was not shared equally among all participants. Several app users discussed their

lack of concern about data security and their comfort when using a mood diary, leaving notes, and answering survey questions:

> *This may sound weird, but I rarely think about data security, to be honest...It is rarely my concern, but I guess it should be.*

Several participants described the potential for collaboration between TransLife users and their therapists and their comfort with sharing mood logs and notes with health providers. Participants linked the possibility of sharing personal information with the benefits of receiving help and enhanced coaching from their therapists:

> *I am currently working with a therapist, and I was really excited to see the insight feature. There were times when I would come to my therapist and tell him that my week was awful. I was rarely able to be more precise than that, and we would spend the whole session trying to dig out my triggers. Now I can bring my diary and just show it to him.*

Ease of Use

Users of TransLife described the app as being easy to use, accessible, simple, and easy to navigate. Participants appreciated the ease of use and convenience of the TransLife app:

> *I don't see how anyone would have issues with navigating this app. It is well-structured and eases you into things.*

The simplicity of the app and its intuitive design was highlighted as aspects likely to promote the continued use of the app. Simplicity, as indicated by a user-friendly and intuitive design, was considered a positive attribute of the app:

> *I believe the app was, in general, very intuitive and easy to use. There was nothing that would confuse me or made me pause and think how to navigate it. This is the main reason I want to use it in the future.*

The onboarding process is a participant's first impression of the app. When it is designed correctly, it increases the likelihood of app adoption. Conversely, a complicated onboarding process may result in users not understanding how to use the app, thus having a negative experience and possibly abandoning it after the first use. Participants mentioned that their onboarding process was easy, simple, and straightforward:

> *I don't remember much from the registration process and setting up my account. That is probably to say that the whole process was easy and painless.*

Trust

Trust is an essential component of mHealth adoption. In general terms, trust can be defined as the willingness of a person in a position of vulnerability to take a risk despite uncertainty, placing confidence in the intentions and competence of another. Users of an mHealth app need to trust the app and its creators to provide meaningful content. To assess participants' comfort with the app and to trust the app with deep thoughts or emotions, we looked through optional free-text notes left together with mood logging. There were many instances when participants shared events from their lives, along with their emotional reactions to these events. For example, one participant described his anxiety about seeing a physician:

> *I'll probably have to go to the doctor tomorrow...I hate going to the doctor as it disrupts my whole day, and I find the experience to be overall stressful, so I'm feeling annoyed at myself for being sick. I hate being sick, especially when it just doesn't go away on its own, and I feel like I have to ask for help and get doctors involved.*

Another participant anxiously anticipated her trip back home after being gone for a long time. She was expecting a negative reaction from her parents to her appearance, and she was mentally preparing to face criticism:

> *Worried about my upcoming trip with my family and feeling very overworked trying to afford plane tickets. I'm hoping to avoid criticism, but it's really inevitable. Maybe things will get better once we have the chance to talk more. It's been a long time.*

Participants shared not only their worries and fears but also their positive emotions, such as this entry from a user who was excited about the upcoming pride weekend:

> *Today is the start of my pride weekend. I just really want to have such a good time. I want to drink a lot and look real cute and run around feeling free and happy...I want to make new friends and run into old ones. I want everyone to like me and want to talk to me and flirt with me and tell me how good I look and how much fun they are having being around me.*

Even more mundane notes, not linked to a significant event such as pride weekends or a trip back home, were still insightful and demonstrated participants' trust and comfort with using the app for self-reflection:

> *A rare night of uninterrupted sleep. Yet I feel pretty tired and emotionally drained this AM. Hoping the sunshine, mild temps, and productive day will boost my mood and energy.*

Satisfaction

Several studies [24,25] link app continuance intention with user satisfaction, defined as "an overall affective response to the gap between prior expectation and perceived performance after consumption." One of the accepted ways to measure user satisfaction is to assess users' willingness to recommend the app to others. Participants were willing to mention this app with their friends:

> *I feel like I've already mentioned this app to friends. Even considering the issue of finding a care provider. This is something that a lot of people are struggling with – finding a friendly and knowledgeable provider. If this app could help them with this while also tracking their moods and connecting them to a wider community – this would be amazing!*

There is a connection between user satisfaction and the perceived value of the app. Most participants viewed the app as useful and connected their satisfaction with TransLife to the benefits of using it:

> *I rated this app so high because I really like the idea of it and I think it is very helpful for a lot of people. It is perfect for finding trends in your moods and for becoming more self-aware.*

Many participants noted the potential for this app to improve and the fact that the new release of TransLife used by a large number of people will make the app especially beneficial. Several features of the app, such as a version of Yelp called local or *question of the day* feature, rely on user-generated content. Participants appreciated the potential for these features to become even more useful and engaging, making the TransLife app even more satisfying to use:

> *I would give this app four out of five because it has room to grow – more resources can be added, more local content. This app has the potential to become a very useful tool for a lot of people I know.*

Discussion

Principal Findings

This study was the first to explore attitudes toward mental health app use among a diverse sample of transgender adults. The purpose was to discover what components or features of a mobile mental health app may resonate with transgender persons and what concerns and suggestions they may have about a mental health app. This addressed an important gap in the literature, as we sought input regarding mental health apps from this priority population with high rates of SI and SAs. The overall findings revealed that transgender people are quite enthusiastic about the role of mobile apps for mental health support and community building. More specifically, we sought to explore the acceptability, use, and feasibility of a new smartphone app intervention, TransLife. Themes that emerged from our analysis suggested that TransLife is an engaging, useful, and acceptable mHealth intervention. Participants reported that the app was easy to use and understand, supported mental self-care, promoted self-awareness, and helped them identify triggers of negative moods.

Acceptability

Mood tracking was well accepted and tolerated by participants. We found that participants wanted to learn about their mood patterns and improve their ability to manage emotional distress more efficiently. Participants were willing to pursue these goals through the regular mood tracking feature, through mood data visualization using the insights feature, through registering contextual information in mood notes, and being willing to share this information with health care providers. The acceptability of this app is enhanced by designing app experience with the transgender population in mind (eg, addressing unique mental health needs) and offering flexible and personalized mood tracking options (eg, integrating contextual information, ability to customize reminders). Acceptability can be further enhanced by increasing app design to be private, discrete, and confidential to use. Participants recognized the potential for mood tracking to facilitate earlier identification of mood changes and enhance clinical care. They also valued the opportunity to view and make sense of their mood data to gain a more individualized and accurate picture of their mental health. This, in turn, presents the opportunity for more dynamic and flexible service provision and monitoring of treatment outcomes.

Usability

Participants thought TransLife was accessible, easy to navigate, and not burdensome. Ease of use is essential in facilitating engagement, and TransLife was noted for being user-friendly and intuitively designed. In addition to regular engagement with the mood logger, many participants chose to write notes in the associated diary without specific prompts, indicating a level of trust in the app. Participants used mood tracking and diary to reflect on their emotions and develop awareness about what may influence positive or negative moods. The question of the day feature sustained engagement and created a community feeling within the app, whereas events and local features connected participants to their local community and helped them identify local resources. Many participants were not engaged in mood tracking daily (4 days per week on average, SD 2.7). When asked to suggest features that would support sustained engagement, participants provided feedback that would be incorporated into the next release of the TransLife. Participants recommended including a transgender-specific news stream, gamification feature, and tailored activity suggestions based on mood (eg, coping strategies).

Although existing studies have highlighted the disparities in mental health outcomes between transgender people and cisgender people, most of this research has examined transgender people as a single group. Given the limitations of this study, we cannot conclude the mental health status and treatment-seeking behavior of various subgroups within the transgender community. However, we noticed several trends among our study participants, suggesting a lack of uniformity in dealing with psychological stressors and using self-help tools. Nonbinary participants reported a previous history of SAs when taking the baseline survey; they made fewer mood entries but often used *anxiousfrustrated* tags, and their engagement with the app was minimal. Transgender men more frequently made use of the free-text entries associated with mood logs, engaged with the app resources more often, and actively suggested app improvements and additional resources. Notes and mood logs of transgender men indicate a higher level of gender dysmorphia. Transgender women frequently reported a history of depression in their baseline surveys and mood logs. During their interviews, transgender women talked about a lack of access to transfriendly mental health services, and they welcomed an app designed with their community in the view that it would equip them with self-care tools for mental health and connect them to transfriendly services.

Limitations

This exploratory pilot study has several limitations. First, the sample size and study design (eg, naturalistic app use and qualitative interviews; n=16), albeit consistent with recommendations for early pilot intervention work [26], does not allow for strong causal conclusions. Second, although this study was useful for gaining insight into the transgender people's experience and perceptions of the app over 2 weeks, it would be valuable to examine whether these perceptions change over a more extended period of naturalistic use. Third, our

participants were highly educated, had stable housing and employment, and lived in urban areas. Therefore, these results may not be generalizable to all transgender populations.

Broader Applicability of Findings

Need for Transgender-Specific Apps That Provide the Community With Resilience-Building Resources

Several mood-logging apps and CBT mHealth interventions have been developed for the general population. These interventions are not tailored for transgender people; they do not address stressors unique to transgender people (eg, ability to log *gender dysmorphia*), and they were not designed to collect longitudinal data. Furthermore, currently, there are no apps designed specifically for transgender people to help them connect with the community of peers, access a variety of community prevented resources, and access transgender-specific health care resources. TransLife is the only app that is tailored to the transgender population and provides them with resources, connects them to their communities, and monitors their psychological well-being.

Need for a More Complex Understanding of Transgender SI Risk

The need to identify modifiable immediate predictors of SI is urgent and widely acknowledged but rarely accomplished because of small sample sizes and methodological limitations. Most studies on transgender SI risk are cross-sectional, collecting data at a single point in time. Although they are able to demonstrate that certain SI risk factors tend to co-occur with reports of past SAs, they are not able to provide evidence that certain risk factors are causally related to earlier behaviors. As SI is associated with SAs, there is a need for prospective data that will allow us to differentiate between risk factors that cause initial SI and risk factors that prolong its duration. The ability to monitor high-risk transgender participants using EMA measurements embedded in TransLife, and the potential to detect certain behavioral or environmental influences and use them as markers of elevated risk may prove to be very meaningful in understanding the short-term risk of SI.

Identification of Modifiable Risk Factors Using EMA

Currently, there have been no studies using EMA to understand short-term modifiable SI and SA risk factors among the transgender population. EMA offers the novel possibility of providing information about the short-term correlates and predictors of SI and SAs, thereby improving our understanding of the suicidal process, detecting persons at risk for SI, and fostering the development of effective strategies for suicide prevention. However, existing studies using EMA mostly focused on negative mood alone while discounting many social, environmental, and behavioral variables. TransLife uses a variety of EMA measures to assess not only negative mood but also behavior, environment, social context, and daily life events.

Conclusions

This is the first study to investigate the acceptability of a mental health app among transgender individuals. The results of this pilot study indicate that TransLife is an engaging, acceptable, and potentially effective mental health intervention. Transgender participants reported many advantages of using TransLife, such as being able to track their mood, connecting to the community, and accessing local resources. This study provides initial support for the acceptability and usability of TransLife as an mHealth intervention designed for the transgender community. The next steps involve further app refinement and more substantial feasibility testing. A large-scale randomized controlled trial is warranted to evaluate the efficacy of TransLife on mental health outcomes among the transgender population.

Acknowledgments

This work was supported by the Center for HIV Identification, Prevention, and Treatment Services (CHIPTS) NIMH grant P30MH058107. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institue of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 15 KB - jmir_v23i3e24023_app1.docx]

References

1. Testa RJ, Michaels MS, Bliss W, Rogers ML, Balsam KF, Joiner T. Suicidal ideation in transgender people: gender minority stress and interpersonal theory factors. J Abnorm Psychol 2017 Jan;126(1):125-136. [doi: 10.1037/abn0000234] [Medline: 27831708]
2. Maguen S, Shipherd JC. Suicide risk among transgender individuals. Psychol Sexual 2010 Mar 31;1(1):34-43. [doi: 10.1080/19419891003634430] [Medline: 30122275]
3. Figueiredo AR, Abreu T. Suicide among LGBT individuals. Eur Psychi 2015 Mar;30:1815. [doi: 10.1016/s0924-9338(15)31398-5]

4. Kleiman EM, Nock MK. Real-time assessment of suicidal thoughts and behaviors. Curr Opin Psychol 2018 Aug;22:33-37. [doi: 10.1016/j.copsyc.2017.07.026] [Medline: 30122275]
5. Kleiman EM, Turner BJ, Fedor S, Beale EE, Huffman JC, Nock MK. Examination of real-time fluctuations in suicidal ideation and its risk factors: Results from two ecological momentary assessment studies. J Abnorm Psychol 2017 Aug;126(6):726-738. [doi: 10.1037/abn0000273] [Medline: 28481571]
6. Meyer IH. Minority stress and mental health in gay men. J Health Soc Behav 1995 Mar;36(1):38-56. [Medline: 7738327]
7. Hendricks ML, Testa RJ. A conceptual framework for clinical work with transgender and gender nonconforming clients: an adaptation of the minority stress model. Professional Psychology: Research and Practice 2012 Oct;43(5):460-467. [doi: 10.1037/a0029597]
8. Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. Ann Behav Med 2005 Aug;30(1):65-73. [doi: 10.1207/s15324796abm3001_8] [Medline: 16097907]
9. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
10. Nuij C, van Ballegooijen W, Ruwaard J, de Beurs D, Mokkenstorm J, van Duijn E, et al. Smartphone-based safety planning and self-monitoring for suicidal patients: rationale and study protocol of the CASPAR (Continuous Assessment for Suicide Prevention And Research) study. Internet Interv 2018 Sep;13:16-23 [FREE Full text] [doi: 10.1016/j.invent.2018.04.005] [Medline: 30206514]
11. Matarazzo BB, Homaifar BY, Wortzel HS. Therapeutic risk management of the suicidal patient: safety planning. J Psychiatr Pract 2014 May;20(3):220-224. [doi: 10.1097/01.pra.0000450321.06612.7a] [Medline: 24847995]
12. Stanley B, Brown G, Brent DA, Wells K, Poling K, Curry J, et al. Cognitive-Behavioral Therapy for Suicide Prevention (CBT-SP): treatment model, feasibility, and acceptability. J Am Acad Child Adolesc Psychiatry 2009 Oct;48(10):1005-1013. [doi: 10.1097/chi.0b013e3181b5dbfe]
13. Stanley B, Brown GK. Safety planning intervention: a brief intervention to mitigate suicide risk. Cogn Behav Pract 2012 May;19(2):256-264. [doi: 10.1016/j.cbpra.2011.01.001]
14. Heckathorn DD. Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. Soc Probl 2002 Feb;49(1):11-34. [doi: 10.1525/sp.2002.49.1.11]
15. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
16. Sauro J. SUPR-Q: a comprehensive measure of the quality of the website user experience. URL: https://uxpajournal.org/supr-q-a-comprehensive-measure-of-the-quality-of-the-website-user-experience/ [accessed 2021-02-24]
17. Nielsen J. Why you only need to test with 5 users. Nielsen Norman Group. 2000. URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ [accessed 2021-02-24]
18. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
19. Lambert SD, Loiselle CG. Combining individual interviews and focus groups to enhance data richness. J Adv Nurs 2008 Apr;62(2):228-237. [doi: 10.1111/j.1365-2648.2007.04559.x] [Medline: 18394035]
20. Tiongson J. Mobile app marketing insights: how consumers really find and use your apps? Think with Google. 2015. URL: https://www.thinkwithgoogle.com/marketing-strategies/app-and-mobile/mobile-app-marketing-insights/ [accessed 2021-02-24]
21. Bowden JL. The process of customer engagement: a conceptual framework. J Mark Theory Pract 2014 Dec 10;17(1):63-74. [doi: 10.2753/mtp1069-6679170105]
22. Vivek SD, Beatty SE, Morgan RM. Customer engagement: exploring customer relationships beyond purchase. J Mark Theory Pract 2014 Dec 08;20(2):122-146. [doi: 10.2753/mtp1069-6679200201]
23. Tarute A, Nikou S, Gatautis R. Mobile application driven consumer engagement. Telematics and Informatics 2017 Jul;34(4):145-156. [doi: 10.1016/j.tele.2017.01.006]
24. Hsiao CH, Chang JJ, Tang KY. Exploring the influential factors in continuance usage of mobile social apps: satisfaction, habit, and customer value perspectives. Telematics and Informatics 2016 May;33(2):342-355. [doi: 10.1016/j.tele.2015.08.014]
25. Bhattacherjee A. Understanding information systems continuance: an expectation-confirmation model. MIS Q 2001 Sep;25(3):351. [doi: 10.2307/3250921]
26. Rounsaville BJ, Carroll KM, Onken LS. A stage model of behavioral therapies research: getting started and moving on from stage I. Hoboken, New Jersey, United States: John Wiley & Sons Inc; 2001:133-142.

Abbreviations

CBT: cognitive behavioral therapy
EMA: ecological momentary assessment
mHealth: mobile health
RDS: respondent-driven sampling
SA: suicide attempt
SI: suicidal ideation

SUPR-Q: Standardized User Experience Percentile Rank Questionnaire
SUS: System Usability Scale

Edited by R Kukafka; submitted 01.09.20; peer-reviewed by R Melia, C Reis; comments to author 21.09.20; revised version received 30.09.20; accepted 17.02.21; published 24.03.21.

Please cite as:
Dubov A, Fraenkel L, Goldstein Z, Arroyo H, McKellar D, Shoptaw S
Development of a Smartphone App to Predict and Improve the Rates of Suicidal Ideation Among Transgender Persons (TransLife): Qualitative Study
J Med Internet Res 2021;23(3):e24023
URL: https://www.jmir.org/2021/3/e24023
doi:10.2196/24023
PMID:33596181

Original Paper

Evaluating Glycemic Control in Patients of South Asian Origin With Type 2 Diabetes Using a Digital Therapeutic Platform: Analysis of Real-World Data

Arjun Krishnakumar[1], MSc, PhD; Ritika Verma[1], MTech, PhD; Rajeev Chawla[2], MD; Aravind Sosale[3], MBBS; Banshi Saboo[4], PhD, MD, DNB; Shilpa Joshi[1], MSc; Maaz Shaikh[1], MBBS, MPH; Abhishek Shah[1], BSc, MBA; Siddhesh Kolwankar[1], MD, MHA, MSMIT; Vinod Mattoo[1], MBBS, MD

[1]Wellthy Therapeutics Pvt Ltd, Mumbai, India

[2]Department of Diabetology, North Delhi Diabetes Centre, New Delhi, India

[3]Department of Diabetology, Diacon Hospital, Bengaluru, India

[4]Department of Diabetology, Dia Care-Diabetes Care and Hormone Clinic, Ahmedabad, India

Corresponding Author:
Siddhesh Kolwankar, MD, MHA, MSMIT
Wellthy Therapeutics Pvt Ltd
20, Nagin Mahal, 5th Floor, 82
Veer Nariman Road, Churchgate
Mumbai, 400020
India
Phone: 91 9833962132
Email: sid.kolwankar@wellthy.care

Abstract

Background: Digital therapeutics are evidence-based therapeutic interventions driven by high-quality software programs for the treatment, prevention, or management of a medical disorder or disease. Many studies in the western population have shown the effectiveness of mobile app–based digital therapeutics for improving glycemic control in patients with type 2 diabetes (T2D). However, few studies have assessed similar outcomes in the South Asian population.

Objective: This study aims to investigate the real-world effectiveness of the Wellthy CARE digital therapeutic for improving glycemic control among the South Asian population of Indian origin.

Methods: We analyzed deidentified data from 102 patients with T2D from India enrolled in a 16-week structured self-management program delivered using the Wellthy CARE mobile app. Patients recorded their meals, weight, physical activity, and blood sugar in the app, and they received lessons on self-care behaviors (healthy eating, being active, monitoring, medication adherence, problem solving, healthy coping, and reducing risks); feedback provided by an artificial intelligence–powered chatbot; and periodic interactions with certified diabetes educators via voice calls and chats. The primary outcome of the program was a change in glycated hemoglobin A_{1c} (HbA_{1c}). Secondary outcomes included the difference between preintervention and postintervention fasting blood glucose (FBG) and postprandial blood glucose (PPBG) levels; changes in BMI and weight at the completion of 16 weeks; and the association between program engagement and the changes in HbA_{1c}, FBG, and PPBG levels.

Results: At the end of 16 weeks, the average change in HbA_{1c} was –0.49% (n=102; 95% CI −0.73 to 0.25; *P*<.001). Of all the patients, 63.7% (65/102) had improved HbA_{1c} levels, with a mean change of −1.16% (n=65; 95% CI −1.40 to −0.92; *P*<.001). The mean preintervention and postintervention FBG levels were 145 mg/dL (n=51; 95% CI 135-155) and 134 mg/dL (n=51; 95% CI 122-146; *P*=.02) and PPBG levels were 188 mg/dL (n=51; 95% CI 172-203) and 166 mg/dL (n=51; 95% CI 153-180; *P*=.03), respectively. The mean changes in BMI and weight were –0.47 kg/m^2 (n=59; 95% CI −0.22 to −0.71; *P*<.001) and –1.32 kg (n=59; 95% CI −0.63 to −2.01; *P*<.001), respectively. There was a stepwise decrease in HbA_{1c}, FBG, and PPBG levels as the program engagement increased. Patients in the highest tertile of program engagement had a significantly higher reduction in HbA_{1c} (−0.84% vs −0.06%; *P*=.02), FBG (−21.4 mg/dL vs −0.18 mg/dL; *P*=.02), and PPBG levels (−22.03 mg/dL vs 2.35 mg/dL; *P*=.002) than those in the lowest tertile.

Conclusions: The use of the Wellthy CARE digital therapeutic for patients with T2D showed a significant reduction in the levels of HbA_{1c}, FBG, and PPBG after 16 weeks. A higher level of participation showed improved glycemic control, suggesting the potential of the Wellthy CARE platform for better management of the disease.

(J Med Internet Res 2021;23(3):e17908) doi:10.2196/17908

KEYWORDS

digital therapeutics; type 2 diabetes; behavior change; diabetes self-management; lifestyle intervention; mobile phone

Introduction

Background

South Asians represent approximately one-fourth of the global population and are at a disproportionately higher risk of being diagnosed with type 2 diabetes (T2D). The global population of patients with diabetes was estimated to be 463 million in 2019 and is projected to reach 578 million by 2030, while that in South East Asia was 88 million in 2019 and is expected to reach 115 million by 2030 [1]. This alarming rise has been attributed to the Asian Indian phenotype and lifestyle changes associated with urbanization and sedentary living [2]. Many studies have reported poor glycemic control in patients with T2D in South Asian countries. [3-5]. In India, the management of T2D is limited by the high burden and early onset of the disease [6]. The Asian phenotype with diabetes has shown marked differences in terms of chemical and biochemical characteristics, such as more significant beta cell dysfunction, higher abdominal adiposity, and a higher susceptibility to developing cardiovascular complications than the Caucasian population [7]. Compared with the general population, diabetes, which is implicated as one of the principal causes of premature heart attacks and death, occurs at 50% higher rates and approximately 5 to 10 years earlier in South Asian patients. Moreover, only a quarter of all South Asian patients achieve their key clinical targets compared with the Caucasian population [5].

Lifestyle management is a fundamental aspect of diabetes care and includes diabetes self-management education and support, medical nutrition therapy, physical activity, smoking cessation counseling, and psychosocial care. Behavioral and lifestyle interventions have been recognized as integral aspects of improving outcomes in these patients. An integrated self-management regimen requires patients to adhere to regular self-glucose monitoring, healthy eating, exercise, and regular physician and specialist visits [8]. Self-monitoring of dietary calorie and fat intake and physical activity are also key strategies [9].

Despite robust evidence that lifestyle interventions in diabetes lead to a reduction in risk factors for cardiovascular diseases [10], self-care behaviors for the management of diabetes among South Asians remain poor because of the lack of knowledge, awareness, and education [11,12]. Moreover, patients solely depend on physicians as a source of information and disease knowledge, which they are unable to provide, as they often lack the time needed to effectively engage patients in self-management behaviors during and between consultation visits [13]. Poor glycemic control leads to earlier incidence and greater severity of diabetes-associated complications, leading to higher morbidity, poor quality of life, and loss of productivity, resulting in increased financial burden [14].

Studies that have examined the effectiveness of diabetes education programs in improving glycemic control in Asian Indians have reported contrasting results depending on the region and intervention [15,16]. One possible explanation for the differences in the results is the cultural appropriateness of these programs. South Asian Indians are a diverse culture with numerous languages and dialects, castes and clans, and cooking styles; therefore, it is imperative to understand the influence of misperceptions, culture, and values on disease management in this population, especially because implementation of culturally appropriate programs leads to a greater reduction in hemoglobin A_{1c} (HbA_{1c}) levels [17]. Traditionally, diabetes education and support have been associated with hospital- or clinic-based practices. However, modern care practices aim to meet these needs through the use of digital interventions that combine monitoring, continuous reinforcement of behavior modification, and personalized therapy using technology [13,18,19].

The widespread adoption of mobile phone technologies in middle- and low-income countries underscores the potential for digital therapeutics to negotiate and overcome the practical roadblocks inherent to conventional physicians and in-clinic–based (eg, face-to-face) interventions and education [18]. Technology-enhanced interventions are cost-effective to enable population management and specialized care accessible to adults with T2D in real time across distant geographies [20-22]. However, little is known about the effect of digital therapeutics in South Asian populations. Therefore, there is a need for approaches that use culturally adapted digital therapeutic interventions specifically for this population [23].

Objectives

On the basis of the role of emerging evidence, we hypothesized that a digital therapeutic with an artificial intelligence (AI)–powered decision support system could enhance multiple behavior patterns (self-monitoring of diet, exercise, weight, and blood glucose) among South Asian patients with T2D. The aim of this study is to assess the real-world effect of the Wellthy CARE digital therapeutics platform on glycemic control (HbA_{1c} and blood sugar levels) and other health outcomes (weight) in patients with T2D after 16 weeks of the program.

Methods

Research Design and Participants

We conducted an analysis of deidentified data for 102 patients with T2D on the Wellthy CARE mobile app. The analysis

involved preintervention and postintervention assessment of HbA_{1c}, fasting blood glucose (FBG), postprandial blood glucose (PPBG), BMI, and weight and program engagement measures in a convenience sample of adults with a self-reported diagnosis of T2D.

Participants were recruited through 2 channels: (1) a network of primary care clinics and diabetes centers through referrals from their treating physicians and (2) a social media campaign seeking participants for a smartphone-based diabetes management program. Patients expressed interest in participating in the program either by (1) calling a telephone number provided on the program recruitment handout or (2) filling out a preregistration form linked to the social media campaign where they voluntarily provided their mobile numbers. Patients who expressed interest in participating in the program received a screening phone call from the program staff members, explaining the program. Those who confirmed their willingness to participate underwent a screening interview (based on inclusion and exclusion criteria) to assess their eligibility. Preexisting T2D status was based on a combination of a self-reported diagnosis and a preintervention HbA_{1c} level of 6.5% or higher. User information was recorded on a web-based form, based on which the participants received a final decision on their eligibility to enroll.

Eligible patients were invited to download the mobile app from the Google Play Store using a unique link sent to them via SMS or text messages. On the mobile app, participants were then instructed to set up their unique account with their phone number and a one-time password sent to them. After the account set up, patients were requested to provide informed consent for their participation in the program and the use of their deidentified data for clinical research purposes.

The inclusion criteria for enrollment were as follows:

- Male or female adult patients aged 18 years or above at the time of enrollment, with a self-reported diagnosis of T2D (HbA_{1c} >6.5%)
- In possession of a personal Android (Google) smartphone with an active internet connection and the person should be comfortable in reading English content on the phone
- No history of a major surgical procedure during the previous 6 months or plan for any major surgical procedure in the next 6 months
- Absence of any medical condition that prevented them from walking for 15 minutes to 30 minutes a day

The exclusion criteria for enrollment were as follows:

- Diagnosis of type 1 diabetes, gestational diabetes mellitus, maturity-onset diabetes of young, or any other forms of diabetes
- Undergoing hemodialysis for chronic kidney disease
- History of any serious heart-related event such as a heart attack or stroke in the past 1 year
- Pregnant, nursing, or planning for a pregnancy in the next 6 months

Participants were not compensated for participation but were enrolled in the program at no cost. Informed consent for using their deidentified data for clinical research was obtained from each participant before enrollment in the program. Participation in the program was voluntary, and refusal to grant consent for the use of their deidentified data for research did not affect the participants' enrollment in the program or the quality of care administered to them. This work involved only secondary analyses of deidentified data; hence, no ethics clearance was obtained. The program did not use any investigational product, and all physical assessments and education were used as usual customary care. No change in treatment was made, and patient care was provided according to the normal clinical standards.

Program

The program was a 16-week structured lifestyle coaching delivered through the Wellthy CARE digital therapeutic consisting of the Wellthy CARE mobile app for Android smartphones, a web portal for health coaches to visualize patient data and communicate with them, and an AI-powered decision support system enabled across the platform (Multimedia Appendix 1). The Wellthy CARE app consisted of a secure messaging center and personal health record files with additional diabetes-related information (eg, laboratory values and details of treating physician) and provided access to a learning library and lesson plans and a logbook to review historical data. The comprehensive lesson plan for the program was based on the American Association of Diabetes Educator's AADE7 self-care behaviors that encouraged patients to acquire skills for better diabetes self-management. The program coached participants along 7 tracks that covered eating healthy, becoming more active, improving self-monitoring, improving medication adherence, problem solving, reducing risk, and healthy coping. The Wellthy CARE app allowed patients to log diabetes self-care data (blood glucose, meals, physical activity, and weight) and diabetes management information such as laboratory reports on a mobile phone. The patient then received automated feedback delivered in real time through a *conversation* experience by an AI-powered chatbot that provided educational, behavioral, and motivational messaging specific to the data entered and in the context of the patient's previous clinical, lifestyle, and behavioral data.

The program adopted a digital persuasion model that focused on improving the patients' motivation, reducing the difficulty in performing a particular task, and then delivering appropriate triggers to the patients to take action. This was done through short culturally relevant content delivered in multiple formats such as informative lessons, videos, and protips along with quizzes and storyboards to reinforce the information and simple tasks to prompt actions. Patients could optionally enroll for challenges that required them to repeat an action multiple times during a specific period to enable skill formation. The program adopted a gamified approach to building skills and rewarded patients with virtual trophies as they completed the lessons, tasks, and challenges. A secure messaging system on the Wellthy CARE app was used by patients to communicate directly with their personal health coach at their own convenience.

Health coaches were *virtual* diabetes educators who regularly reviewed patient data, providing personalized feedback during each interaction, and also responded to patient queries. All

patients received a voice call at the start of the program. Health coaches could supplement automated messages with electronic messages sent to patients via a secure messaging system. Health coach messages were based on longitudinal data trends, providing weekly and monthly summaries to patients on their performance. About 50% of patients made voice calls to or received voice calls from the health coach during the program, in addition to communicating over the messaging system. This was an incremental and supportive program for the existing standard of care recommended by the treating physician.

Measures

A total of 102 participants completed and logged data during the 16 weeks on the Wellthy CARE platform. The participants self-reported their age, gender, height, and weight in the Wellthy CARE mobile app. All patients received a Gluco One (Dr Morepen) glucose meter and strips. The HbA_{1c} test was performed before and after the completion of the intervention by an independent pathological laboratory that reported the values directly.

Outcomes

The primary outcome of this study was the change in HbA_{1c} level at the completion of the program (16 weeks). Secondary objectives included the difference between the mean preintervention and postintervention FBG and PPBG levels in participants who reported more than one blood sugar reading; change in BMI and weight in participants that reported more than one weight log; and association between program engagement, measured as the total number of interactions with the health coach and the AI-powered chatbot, and the change in HbA_{1c}, FBG, and PPBG levels. In addition, we assessed the differences among patients who had an improvement of ≥0.4% in HbA_{1c} (responders) and those who did not (nonresponders).

Statistical Analysis

Statistical analysis was performed using R software (version 3.4.3; the R Foundation). The results were analyzed to identify program starters (ie, those who completed at least one skill) and program completers (ie, those who completed at least 6 weeks of lessons or continued with health coaching through the duration of the program). Baseline characteristics were compared between subgroups using chi-square tests or Fisher exact test for categorical variables and 2-sample two-tailed *t* tests analysis for continuous variables and paired *t* tests for comparing preintervention and postintervention values for continuous variables. *P* values <.05 were considered statistically significant.

Results

Baseline Characteristics of Patients

Figure 1 shows the data for participation and retention during the 16-week program. A total of 102 patients from 18 cities in India participated in the study and completed the program. The baseline characteristics of the participants are summarized in Table 1. There were no statistically significant differences in the baseline characteristics between male and female participants. The average duration of time spent with the personal health coach was 106 minutes (n=87; 95% CI 65-147) over the 16-week period, whereas that with the AI-powered chatbot was 88 minutes (n=102; 95% CI 66-110).

Figure 1. Participant recruitment and retention flowchart.

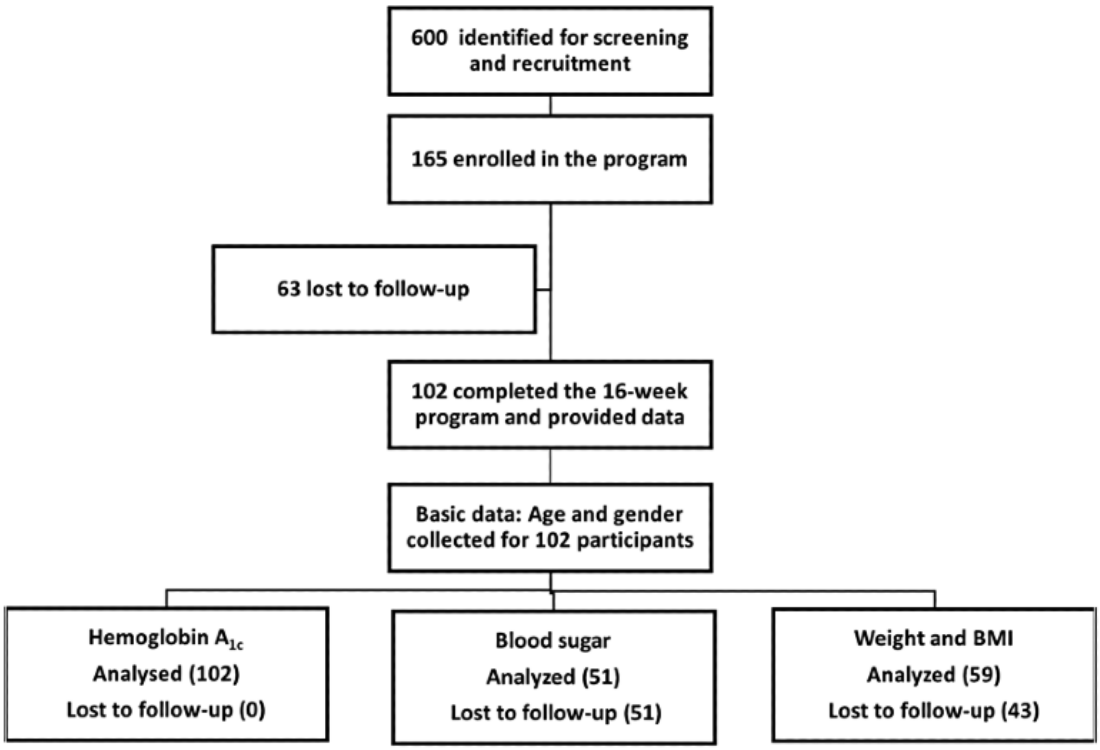


Table 1. Patient characteristics at the start of the program. All CI values are at 95% significance level. *P* value has been shown for comparison of male and female data.

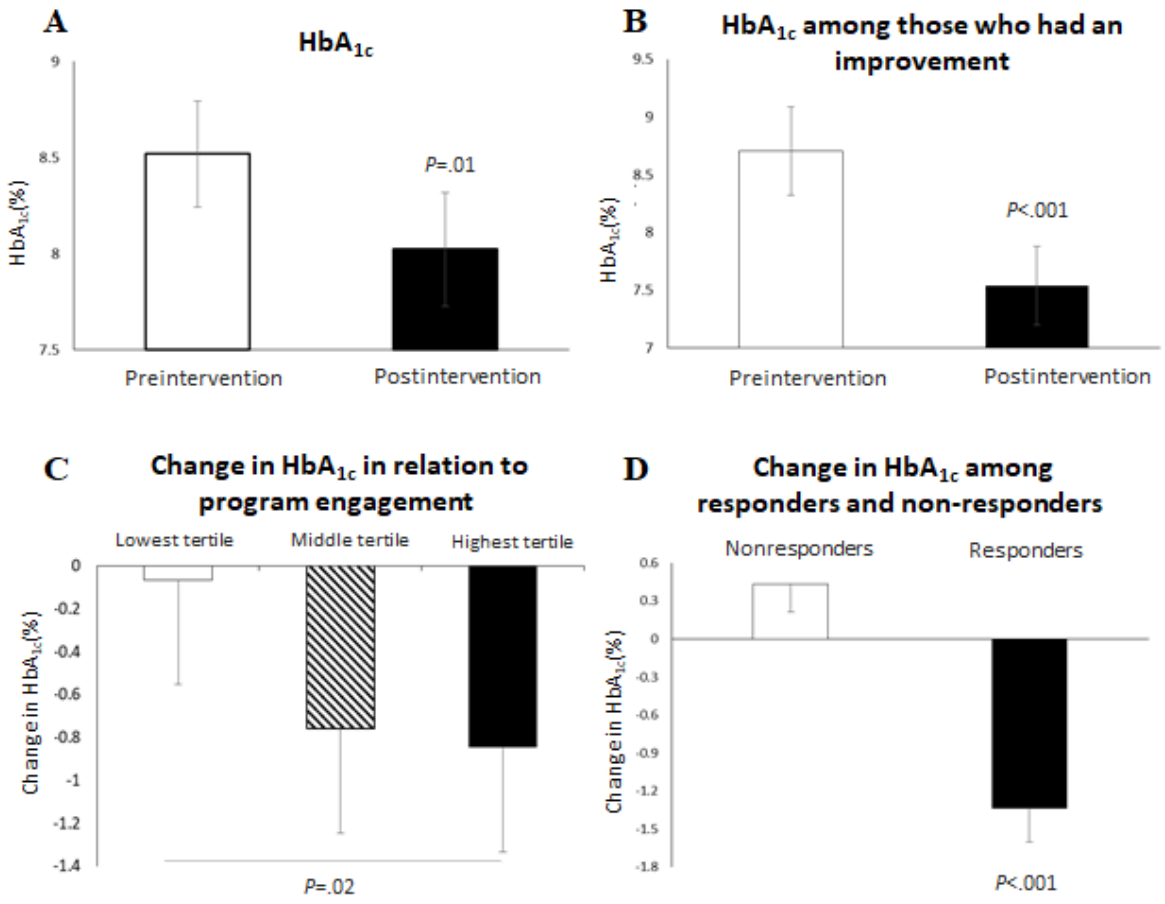
Characteristics assessed	Overall	Male patients	Female patients	*P* value
Gender, n (%)	102 (100)	70 (68.6)	32 (31.4)	—[a]
Age (years), mean (CI)	50.8 (49.2-52.4)	51.4 (49.40-53.4)	49.7 (46.8-52.5)	.32
Weight (kg), mean (CI)	77.3 (80.5-74.0)	78.4 (82.2-74.7)	74.6 (81.1-68.1)	.31
BMI (kg/m^2), mean (CI)	28.1 (29.2-26.4)	27.4 (28.7-26.1)	29.7 (31.8-27.6)	.07
Baseline hemoglobin A_{1c} (%), mean (CI)	8.5 (8.2-8.8)	8.5 (8.8-8.2)	8.6 (9.1-8.0)	.71

[a]Statistical comparison not applicable.

Change in HbA_{1c}

The mean change in HbA_{1c} among all patients was −0.49% (n=102; 95% CI −0.73 to −0.25; *P*<.001; Figure 2). Of all the patients, 63.7% (65/102) had improved HbA_{1c} levels with a mean change of −1.16% (95% CI −1.40 to −0.92; *P*<.001; Figure 2). Among the patients that had an improvement in HbA_{1c} levels, 48.0% (49/102) had a decrease of 0.5% or more and 27.5% (28/102) had a decrease of 1% or more; moreover, 29.4% (30) of the patients had a follow-up HbA_{1c} level less than or equal to 7%. Among those with a baseline HbA_{1c} level ≥7.5%, the mean change was −0.57% (n=59; 95% CI −0.90 to −0.25; *P*<.001).

Figure 2. Comparison of HbA_{1c} levels. (A) Mean change in HbA_{1c} among all participants, (B) mean change in HbA_{1c} among those who had improvements in HbA_{1c} levels, (C) relationship between program engagement and change in HbA_{1c} levels, and (D) mean change in HbA_{1c} among responders and nonresponders. Error bars represent confidence intervals. HbA_{1c}: hemoglobin A_{1c}.



The relationship between program engagement and changes in HbA_{1c} levels was assessed after 16 weeks of the program. There was a stepwise decrease in HbA_{1c} levels as the program engagement level increased (Figure 2). Among those with a baseline HbA_{1c}>7.5%, the lowest tertile of mobile app engagers reduced their HbA_{1c} levels by 0.06% (95% CI −0.56 to 0.43) and the middle tertile of program engagers reduced their HbA_{1c} levels by 0.76% (95% CI −1.33 to −0.18); however, those in

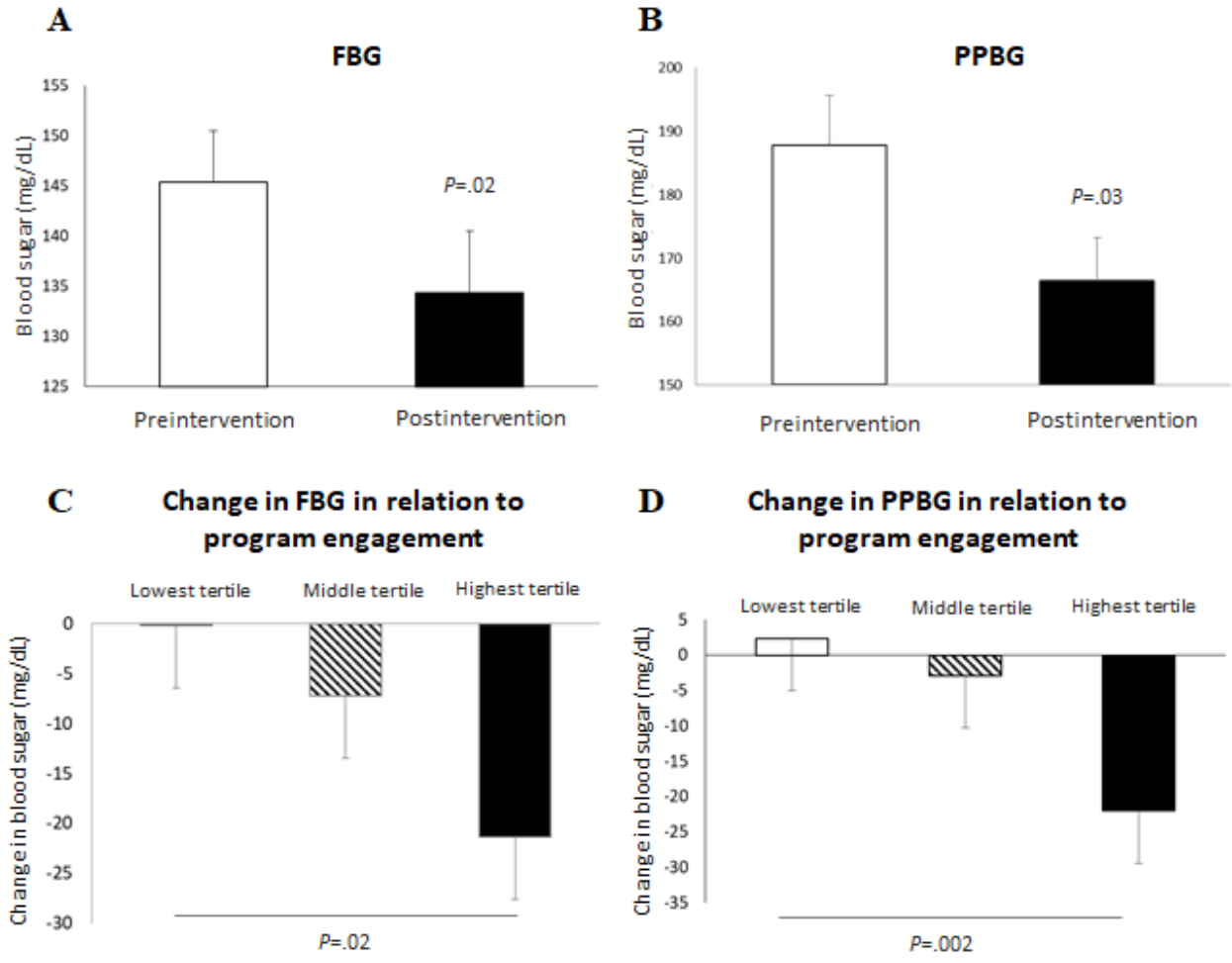
the highest tertile of program engagers significantly reduced their HbA_{1c} levels by 0.84% as compared with patients in lowest tertile (95% CI –1.33 to –0.35; lowest vs highest tertile, *P*=.02).

Responders were grouped as patients who showed an improvement of ≥0.4% in HbA_{1c} levels and those who did not were grouped as nonresponders. The mean change in HbA_{1c} among responders (51/102, 50%) was −1.34% (95% CI −1.07% to −1.60%), and that among nonresponders (51/102, 50%) was 0.43% (95% CI 0.65%-0.21%; *P*<.001; Figure 2).

Change in Blood Sugar

Among patients who reported more than one blood sugar reading (51/102), there was a significant difference between the mean preintervention and postintervention FBG (145 mg/dL; 95% CI 135-155 vs 134 mg/dL; 95% CI 122-146; *P*=.02) and PPBG (188 mg/dL; 95% CI 172-203 vs 166 mg/dL; 95% CI 153-180; *P*=.03) values (Figure 3).

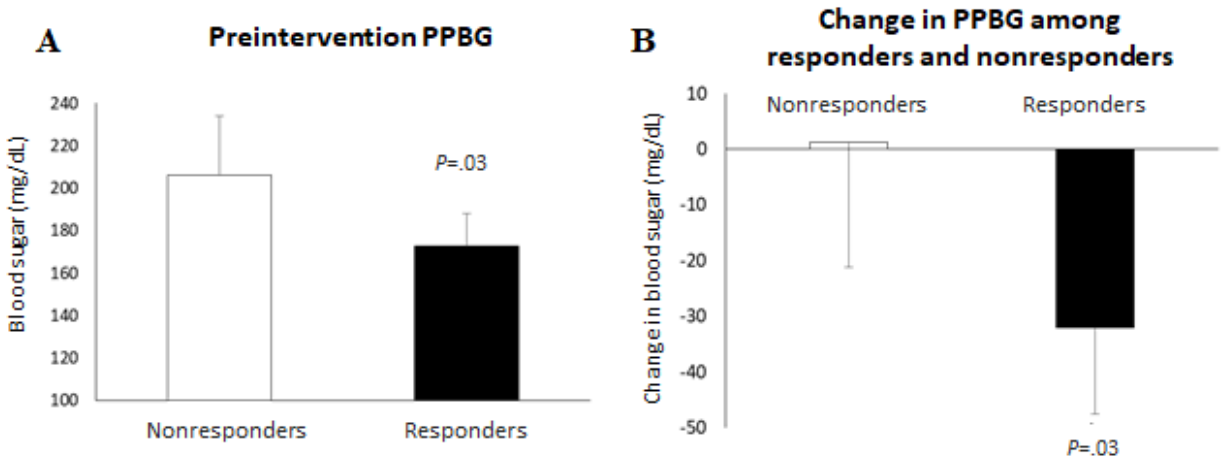
Figure 3. Change in blood sugar levels. Difference between (A) mean FBG levels and (B) mean PPBG levels reported in the first week and final week of the intervention. Relationship between mobile app use and change in (C) FBG and (D) PPBG levels. Error bars represent confidence intervals. FBG: fasting blood glucose; PPBG: postprandial blood glucose.



There was a stepwise decrease in FBG and PPBG levels as the mobile app engagement level increased. The lowest tertile of app engagers reduced their FBG by 0.18 mg/dL and there was an increase in PPBG by 2.35 mg/dL; the middle tertile of app engagers reduced their FBG and PPBG by 7.25 mg/dL and 2.84 mg/dL, respectively; and those in the highest tertile of app engagers reduced their FBG by 21.4 mg/dL (*P*=.02, highest vs lowest) and PPBG by 22.03 mg/dL (*P*=.02, highest vs middle; *P*=.002, highest vs lowest; Figure 3).

Responders had a significantly lower mean preintervention PPBG of 172 mg/dL (n=28; 95% CI 188-157) than the mean preintervention PPBG of 206 mg/dL among nonresponders (n=23, nonresponders; 95% CI 234-178; *P*=.03; Figure 4). The responders reduced their PPBG by −32 mg/dL (n=21; 95% CI −16.4 to −47.2) as compared with 1.25 mg/dL (n=20; 95% CI 23.77 to −21.27; *P*=.03) among the nonresponders (Figure 4).

Figure 4. (A) Difference in the mean preintervention PPBG between nonresponders and responders. (B) Difference in the mean change in PPBG between the nonresponders and responders. PPBG: postprandial blood glucose.

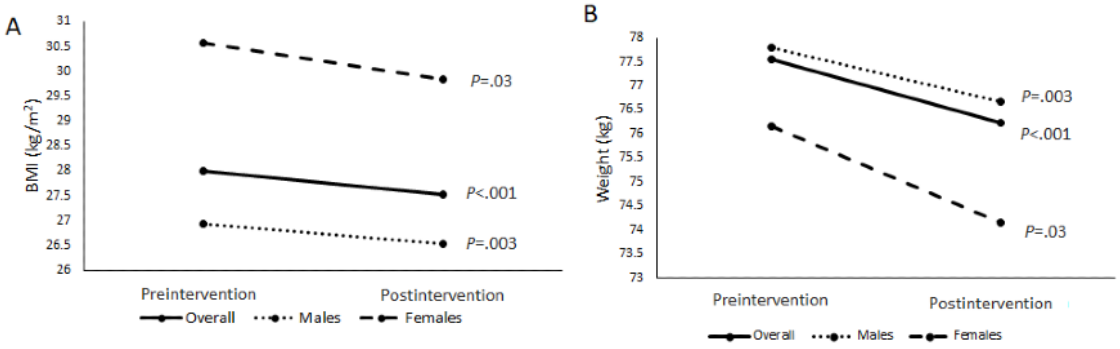


Change in BMI and Weight

There was a significant change in BMI (Figure 5) and in weight (Figure 5). The change in the mean BMI was −0.47 kg/m^2 (n=59; 95% CI −0.22 to −0.71; *P*<.001), and the mean BMI reduction in males and females was −0.38 kg/m^2 (n=42; 95% CI −0.14 to −0.62; *P*=.003) and −0.73 kg/m^2 (n=17; 95% CI −0.12 to −1.34; *P*=.03), respectively. The change in mean weight was −1.32 kg (n=59; 95% CI −0.63 to −2.01; *P*<.001), and the mean weight change in males and females was −1.11 kg (n=42; 95% CI −0.40 to −1.83; *P*=.003) and −2.00 kg (n=17; 95% CI −0.38 to −3.62; *P*=.03), respectively.

The responders reduced their weight by −2.36 kg (n=27; 95% CI −1.32 to −3.41) as compared with −0.26 kg (n=23; 95% CI 0.42 to −0.95; *P*=.005) among the nonresponders. There was no significant difference in the mean preintervention weight between responders and nonresponders.

Figure 5. Change in BMI and weight. Difference between the (A) mean BMI and (B) mean weight reported in the first week and final week of the intervention.



Discussion

Principal Findings

This study assessed the effectiveness of the Wellthy CARE digital therapeutics platform to improve glycemic control (reduction in HbA_{1c} levels and blood sugar levels) and reduce weight in patients of South Asian origin with T2D after a duration of 16 weeks of the program. Patients who used the digital therapeutic for 16 weeks reduced their HbA_{1c} levels by 0.49% (n=102), FBG by 11 mg/dL (n=51), PPBG by 21 mg/dL (n=51), weight by 1.32 kg (n=59), and BMI by 0.47 kg/m^2 (n=59). Patients with the highest engagement significantly reduced their HbA_{1c} levels by 0.84%, FBG by 21.4 mg/dL, and PPBG by 22.03 mg/dL. There was a higher reduction in HbA_{1c}, FBG, and PPBG levels as program engagement increased.

The effectiveness of the digital therapeutic was suggested by the significant reduction in HbA_{1c} levels. A 0.5% to 1% change in HbA_{1c} level is considered clinically significant to reduce the risk of comorbid conditions; even the US Food and Drug Administration requires a 0.4% change in HbA_{1c} level for drug evaluations [24]. The results of the United Kingdom Prospective Diabetes Study indicated that a 0.9% decrease in HbA_{1c} level was associated with a 25% reduction in microvascular complications, a 10% decrease in diabetes-related mortality, and a 6% reduction in all-cause mortality [25,26]. Hence, the average reduction of 1.34% in HbA_{1c} among the responders is significant in reducing the risk of complications and mortality.

FBG and PPBG levels are indicators of glycemic control and both are correlated with HbA_{1c} [27]. PPBG level has been shown to predict cardiovascular risk and all-cause mortality and has been shown to have a stronger correlation with HbA_{1c}. There was a reduction in PPBG levels among responders by 32 mg/dL,

which indicates an additional benefit of reduction in cardiovascular risk among these patients.

Weight reduction is one of the goals in the management of diabetes. Weight is associated with an increased risk of cardiovascular diseases [28], and a reduction in weight is associated with improvements in HbA_{1c} levels [29]. Although there was only a modest reduction of 1.32 kg weight among the participants after 16 weeks of the program, the significant reduction in weight among the responders (2.36 kg) highlights the importance of reducing weight in improving HbA_{1c} levels. It was also observed that females showed a higher reduction in mean BMI than males, even when no statistically significant difference between the BMI of the 2 groups was present at the preintervention stage. There could be several reasons for this, including differences in metabolism or hormonal balance between the 2 groups [30,31]. A substantial reason for such an effect could not be analyzed in this analysis.

To our knowledge, this work is one of the first to report the real-world effectiveness of implementing a digital therapeutic that is a structured, behavioral, and self-management program, which is augmented and delivered using a mobile-enabled app and an AI-powered decision support system for South Asian patients with T2D. Furthermore, the study highlighted the effect of the digital therapeutic to improve self-management of diabetes by demonstrating improvement in HbA_{1c}, FBG, PPBG levels and reduction in BMI and weight. The real-world feasibility and acceptance were demonstrated by the observation that most of the users entered their HbA_{1c} readings during the program.

In recent years, considerable evidence has been generated about the effect of digital therapeutics in effectively promoting self-care and health behavior changes among patients with chronic diseases and mental health issues [32,33]. In countries such as India that lack appropriate diabetes self-management programs and where there is an overdependence for disease knowledge and support on physicians, the need for appropriate digital interventions is even more pronounced. Moreover, it has been reported that diabetes management interventions targeted at South Asian patients are heterogeneous, yielding variable and limited success in reducing HbA_{1c} levels [34]. Geographic and cultural diversity are significant barriers to the adoption of new treatments and technologies. However, the results here are a vindication of the value of a digital therapeutic such as Wellthy CARE across a wide range of population types and suggest that digital intervention provides a reasonable alternative in not only plugging some of the resourcing gaps in underserved countries but also has the potential to provide meaningful improvements in diabetes management.

Limitations

One of the limitations of this study is the lack of a control group, as real-world data were retrospectively analyzed. Other limitations include selection bias due to multiple approaches (physician-recommended and voluntary approach) used for the selection of participants and reliance on self-reported disease biomarkers. The program was conducted for a short duration, and we did not independently quantify the effect of other behavioral and lifestyle measures on glycemic outcomes. This analysis can prove to be beneficial in providing important insights, especially when comparing the results of the Wellthy CARE platform with other digital therapeutics platforms. Real-world data helped in the study of people outside a controlled set of conditions. This increased the variability in the study; however, such studies have shown to be very helpful for the evaluation of new technologies. A variation in the number of male and female participants was noted, which was an undesirable bias due to social, economic, or other factors. The loss of data during follow-up also limited the scope of the study. However, the reporting for HbA_{1c} values was substantially high, with 62% of the enrolled users logging their preintervention and postintervention values during the entire program. This study showed significant improvement in HbA_{1c} and blood sugar values after a 16-week program, suggesting the potential of the Wellthy CARE platform for improving glycemic control in patients with T2D. However, future studies with a larger sample with better control will be able to further establish the effectiveness of the program.

Strengths

The strength of this work was its design, which enabled us to closely simulate real-world implementations. The effectiveness of the intervention in documenting meaningful changes in HbA_{1c} levels, positive relationship of mobile app engagement with improvement in glycemic control, good rate of retention, and successful data collection contributed to its strength. The other strengths were that the app was implemented with the support of the same developmental processes and core support team that were involved in the development phase.

Conclusions

This work is one of the first to report the real-world effectiveness of a digital therapeutic that was a structured, behavioral, and self-management intervention delivered using a mobile-enabled app and an AI-powered decision support system—Wellthy CARE in improving glycemic control among South Asians with T2D. The intervention demonstrated incremental reduction in HbA_{1c}, FBG, and PPBG values with greater levels of engagement, suggesting the feasibility, acceptance, and value of using a digital therapeutic in resource-constrained countries of South Asia. The study findings can be explored further to evaluate the long-term acceptability, cost-effectiveness, and durability of the principal findings and its feasibility to be applied to a larger, culturally similar population.

Acknowledgments

The authors would like to thank all the participants who participated in the program and the Research Society for the Study of Diabetes in India for their support.

Conflicts of Interest

AS is the chief executive officer and a shareholder of Wellthy Therapeutics Pvt Ltd. MS and AK are former employees and shareholders in Wellthy Therapeutics Pvt Ltd. SK is a current employee of Wellthy Therapeutics Pvt Ltd. RV is a current consultant in Wellthy Therapeutics Pvt Ltd. VM and SJ are former consultants of Wellthy Therapeutics Pvt Ltd.

Multimedia Appendix 1
Schematic of the Wellthy CARE digital therapeutics platform.
[PNG File , 4493 KB - jmir_v23i3e17908_app1.png]

References

1. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, IDF Diabetes Atlas Committee. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9 edition. Diabetes Res Clin Pract 2019 Nov;157. [doi: 10.1016/j.diabres.2019.107843] [Medline: 31518657]
2. Unnikrishnan R, Anjana RM, Mohan V. Diabetes in South Asians: is the phenotype different? Diabetes 2014 Jan;63(1):53-55 [FREE Full text] [doi: 10.2337/db13-1592] [Medline: 24357697]
3. Aschner P, Gagliardino JJ, Ilkova H, Lavalle F, Ramachandran A, Mbanya JC, et al. Persistent poor glycaemic control in individuals with type 2 diabetes in developing countries: 12 years of real-world evidence of the International Diabetes Management Practices Study (IDMPS). Diabetologia 2020 Apr;63(4):711-721 [FREE Full text] [doi: 10.1007/s00125-019-05078-3] [Medline: 31901950]
4. Ramachandran A, Jain SM, Mukherjee S, Phatak S, Pitale S, Singh SK, et al. Suboptimal glycemic control among subjects with diabetes mellitus in India: a subset analysis of cross-sectional wave-7 (2016) data from the International Diabetes Management Practices Study (IDMPS). Ther Adv Endocrinol Metab 2020;11 [FREE Full text] [doi: 10.1177/2042018820937217] [Medline: 32647562]
5. Sohal T, Sohal P, King-Shier KM, Khan NA. Barriers and facilitators for Type-2 Diabetes Management in South Asians: a systematic review. PLoS One 2015;10(9) [FREE Full text] [doi: 10.1371/journal.pone.0136202] [Medline: 26383535]
6. Anjana RM, Shanthi Rani CS, Deepa M, Pradeepa R, Sudha VL, Nair DH, et al. Incidence of diabetes and prediabetes and predictors of progression among Asian Indians: 10-Year follow-up of the Chennai Urban Rural Epidemiology Study (CURES). Diabetes Care 2015 Aug;38(8):1441-1448. [doi: 10.2337/dc14-2814] [Medline: 25906786]
7. Lim LL, Tan ATB, Moses K, Rajadhyaksha V, Chan SP. Place of sodium-glucose cotransporter-2 inhibitors in East Asian subjects with type 2 diabetes mellitus: insights into the management of Asian phenotype. J Diabetes Complications 2017 Feb;31(2):494-503. [doi: 10.1016/j.jdiacomp.2016.10.008] [Medline: 27866701]
8. American Diabetes Association. 5. Lifestyle management: standards of medical care in diabetes-2019. Diabetes Care 2019 Jan 17;42(Suppl 1):S46-S60. [doi: 10.2337/dc19-S005] [Medline: 30559231]
9. Wang J, Cai C, Padhye N, Orlander P, Zare M. A behavioral lifestyle intervention enhanced with multiple-behavior self-monitoring using mobile and connected tools for underserved individuals with type 2 diabetes and comorbid overweight or obesity: pilot comparative effectiveness trial. JMIR Mhealth Uhealth 2018 Apr 10;6(4):e92 [FREE Full text] [doi: 10.2196/mhealth.4478] [Medline: 29636320]
10. Chen L, Pei J, Kuang J, Chen H, Chen Z, Li Z, et al. Effect of lifestyle intervention in patients with type 2 diabetes: a meta-analysis. Metabolism 2015 Feb;64(2):338-347 [FREE Full text] [doi: 10.1016/j.metabol.2014.10.018] [Medline: 25467842]
11. Mohandas A, Bhasin SK, Upadhyay M, Madhu SV. Diabetes self care activities among adults 20 years and above residing in a resettlement colony in East Delhi. Indian J Public Health 2018;62(2):104-110 [FREE Full text] [doi: 10.4103/ijph.IJPH_249_17] [Medline: 29923533]
12. Rajasekharan D, Kulkarni V, Unnikrishnan B, Kumar N, Holla R, Thapar R. Self-care activities among patients with diabetes attending a tertiary care hospital in Mangalore Karnataka, India. Ann Med Health Sci Res 2015;5(1):59-64 [FREE Full text] [doi: 10.4103/2141-9248.149791] [Medline: 25745579]
13. Burke SD, Sherr D, Lipman RD. Partnering with diabetes educators to improve patient outcomes. Diabetes Metab Syndr Obes 2014;7:45-53 [FREE Full text] [doi: 10.2147/DMSO.S40036] [Medline: 24550679]
14. Joshi SR. Diabetes care in India. Ann Glob Health 2015;81(6):830-838 [FREE Full text] [doi: 10.1016/j.aogh.2016.01.002] [Medline: 27108150]
15. Bellary S, O'Hare JP, Raymond NT, Gumber A, Mughal S, Szczepura A, UKADS Study Group. Enhanced diabetes care to patients of South Asian ethnic origin (the United Kingdom Asian Diabetes Study): a cluster randomised controlled trial. Lancet 2008 May 24;371(9626):1769-1776. [doi: 10.1016/S0140-6736(08)60764-3] [Medline: 18502301]

16. Doupis J, Festas G, Tsilivigos C, Efthymiou V, Kokkinos A. Smartphone-based technology in diabetes management. Diabetes Ther 2020 Mar;11(3):607-619 [FREE Full text] [doi: 10.1007/s13300-020-00768-3] [Medline: 31983028]
17. Bhurji N, Javer J, Gasevic D, Khan NA. Improving management of type 2 diabetes in South Asian patients: a systematic review of intervention studies. BMJ Open 2016 Apr 20;6(4) [FREE Full text] [doi: 10.1136/bmjopen-2015-008986] [Medline: 27098819]
18. Kvedar JC, Fogel AL, Elenko E, Zohar D. Digital medicine's march on chronic disease. Nat Biotechnol 2016 Mar;34(3):239-246. [doi: 10.1038/nbt.3495] [Medline: 26963544]
19. Sepah SC, Jiang L, Peters AL. Long-term outcomes of a web-based diabetes prevention program: 2-year results of a single-arm longitudinal study. J Med Internet Res 2015 Apr 10;17(4):e92 [FREE Full text] [doi: 10.2196/jmir.4052] [Medline: 25863515]
20. Mohan V, Deepa R, Deepa M, Somannavar S, Datta M. A simplified Indian Diabetes Risk Score for screening for undiagnosed diabetic subjects. J Assoc Physicians India 2005 Sep;53:759-763. [Medline: 16334618]
21. Zhang L, He X, Shen Y, Yu H, Pan J, Zhu W, et al. Effectiveness of smartphone app-based interactive management on glycemic control in Chinese patients with poorly controlled diabetes: randomized controlled trial. J Med Internet Res 2019 Dec 09;21(12) [FREE Full text] [doi: 10.2196/15401] [Medline: 31815677]
22. Idris I, Hampton J, Moncrieff F, Whitman M. Effectiveness of a digital lifestyle change program in obese and type 2 diabetes populations: service evaluation of real-world data. JMIR Diabetes 2020 Jan 20;5(1) [FREE Full text] [doi: 10.2196/15189] [Medline: 31958064]
23. Joseph LM, Berry D, Jessup A. Management of type 2 diabetes in Asian Indians: a review of the literature. Clin Nurs Res 2015 Apr;24(2):188-210. [doi: 10.1177/1054773814531600] [Medline: 24789938]
24. American Diabetes Association. 6. Glycemic targets: standards of medical care in diabetes—2019. Dia Care 2018 Dec 17;42(Supplement 1):S61-S70. [doi: 10.2337/dc19-s006]
25. Chrvala CA, Sherr D, Lipman RD. Diabetes self-management education for adults with type 2 diabetes mellitus: a systematic review of the effect on glycemic control. Patient Educ Couns 2016 Jun;99(6):926-943 [FREE Full text] [doi: 10.1016/j.pec.2015.11.003] [Medline: 26658704]
26. Katulanda P, Jayawardena R, Jayawardana R, Ranasinghe P, Rezvi Sheriff MH, Matthews DR. Physical activity patterns and correlates among adults from a developing country: the Sri Lanka Diabetes and Cardiovascular Study. Public Health Nutr 2013 Sep;16(9):1684-1692. [doi: 10.1017/S1368980012003990] [Medline: 22995708]
27. Ketema EB, Kibret KT. Correlation of fasting and postprandial plasma glucose with HbA1c in assessing glycemic control; systematic review and meta-analysis. Arch Public Health 2015;73:43 [FREE Full text] [doi: 10.1186/s13690-015-0088-6] [Medline: 26413295]
28. Davies MJ, D'Alessio DA, Fradkin J, Kernan WN, Mathieu C, Mingrone G, et al. Management of hyperglycemia in Type 2 Diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care 2018 Dec;41(12):2669-2701 [FREE Full text] [doi: 10.2337/dci18-0033] [Medline: 30291106]
29. Gummesson A, Nyman E, Knutsson M, Karpefors M. Effect of weight reduction on glycated haemoglobin in weight loss trials in patients with type 2 diabetes. Diabetes Obes Metab 2017 Sep;19(9):1295-1305. [doi: 10.1111/dom.12971] [Medline: 28417575]
30. Chan Z, Chooi YC, Ding C, Choo J, Sadananthan SA, Michael N, et al. Sex differences in glucose and fatty acid metabolism in Asians who are nonobese. J Clin Endocrinol Metab 2019 Jan 01;104(1):127-136. [doi: 10.1210/jc.2018-01421] [Medline: 30252100]
31. Tramunt B, Smati S, Grandgeorge N, Lenfant F, Arnal J, Montagner A, et al. Sex differences in metabolic regulation and diabetes susceptibility. Diabetologia 2020 Mar;63(3):453-461 [FREE Full text] [doi: 10.1007/s00125-019-05040-3] [Medline: 31754750]
32. Alkhaldi G, Modrow K, Hamilton F, Pal K, Ross J, Murray E. Promoting engagement with a Digital Health Intervention (HeLP-Diabetes) using email and text message prompts: mixed-methods study. Interact J Med Res 2017 Aug 22;6(2):e14 [FREE Full text] [doi: 10.2196/ijmr.6952] [Medline: 28829328]
33. Lee JA, Choi M, Lee SA, Jiang N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: a systematic review. BMC Med Inform Decis Mak 2018 Feb 20;18(1):12 [FREE Full text] [doi: 10.1186/s12911-018-0591-0] [Medline: 29458358]
34. Ballotari P, Ferrari F, Ballini L, Chiarenza A, Manicardi V, Giorgi Rossi P. Lifestyle-tailored interventions for South Asians with type 2 diabetes living in high-income countries: a systematic review. Acta Diabetol 2017 Aug;54(8):785-794. [doi: 10.1007/s00592-017-1008-8] [Medline: 28585044]

Abbreviations

AI: artificial intelligence
FBG: fasting blood glucose
HbA_{1c}: hemoglobin A_{1c}

PPBG: postprandial blood glucose
T2D: type 2 diabetes

Edited by G Eysenbach, R Kukafka; submitted 21.01.20; peer-reviewed by J Santos, M Peeples, K Fitzner, J Kvedar; comments to author 10.08.20; revised version received 25.09.20; accepted 22.01.21; published 25.03.21.

Please cite as:
Krishnakumar A, Verma R, Chawla R, Sosale A, Saboo B, Joshi S, Shaikh M, Shah A, Kolwankar S, Mattoo V
Evaluating Glycemic Control in Patients of South Asian Origin With Type 2 Diabetes Using a Digital Therapeutic Platform: Analysis of Real-World Data
J Med Internet Res 2021;23(3):e17908
URL: https://www.jmir.org/2021/3/e17908
doi:10.2196/17908
PMID:33764306

Original Paper

Public Interest in Cosmetic Surgical and Minimally Invasive Plastic Procedures During the COVID-19 Pandemic: Infodemiology Study of Twitter Data

Wenhui Liu[1*], MD, PhD; Zhiru Wei[1*], MD, PhD; Xu Cheng[1], MD; Ran Pang[1], MD; Han Zhang[1], MD; Guangshuai Li[1], MD, PhD

Plastic and Reconstructive Surgery, First Affiliated Hospital, Zhengzhou University, Zhengzhou, China

*these authors contributed equally

Corresponding Author:
Guangshuai Li, MD, PhD
Plastic and Reconstructive Surgery
First Affiliated Hospital
Zhengzhou University
1 Jianshe E Road
Erqi District
Zhengzhou,
China
Phone: 86 0371 66862241
Email: liguangshuai@zzu.edu.cn

Abstract

Background: The unprecedented COVID-19 pandemic has brought drastic changes to the field of plastic surgery. It is critical for stakeholders in this field to identify the changes in public interest in plastic procedures to be adequately prepared to meet the challenges of the pandemic.

Objective: The aim of this study is to examine tweets related to the public interest in plastic procedures during the COVID-19 pandemic and to help stakeholders in the field of plastic surgery adjust their practices and sustain their operations during the current difficult situation of the pandemic.

Methods: Using a web crawler, 73,963 publicly accessible tweets about the most common cosmetic surgical and minimally invasive plastic procedures were collected. The tweets were grouped into three phases, and the tweeting frequencies and Google Trends indices were examined. Tweeting frequency, sentiment, and word frequency analyses were performed with Python modules.

Results: Tweeting frequency increased by 24.0% in phase 2 and decreased by 9.1% in phase 3. Tweets about breast augmentation, liposuction, and abdominoplasty ("tummy tuck") procedures consecutively increased over the three phases of the pandemic. Interest in Botox and chemical peel procedures revived first when the lockdown was lifted. The COVID-19 pandemic was associated with a negative impact on public sentiment about plastic procedures. The word frequency pattern significantly changed after phase 1 and then remained relatively stable.

Conclusions: According to Twitter data, the public maintained their interest in plastic procedures during the COVID-19 pandemic. Stakeholders should consider refocusing on breast augmentation, liposuction, and abdominoplasty procedures during the current phase of the pandemic. In the case of a second wave of COVID-19, stakeholders should prepare for a temporary surge of Botox and chemical peel procedures.

(J Med Internet Res 2021;23(3):e23970) doi:10.2196/23970

KEYWORDS

COVID-19; Twitter; Google Trends; plastic procedure; trend; survey; surgery; social media

Introduction

The unprecedented COVID-19 pandemic has brought drastic changes not only to daily life but also to the field of plastic surgery [1-4]. Due to repeated periods of quarantine and the deteriorating economy, many people are not scheduling nonessential medical services, such as plastic procedures; this has resulted in concerning prospects for stakeholders in this field, including plastic surgeons, hospital administrators, and clinic owners. These stakeholders require a better understanding of the public interest in common plastic procedures to adjust their practices so that they can sustain operations during this difficult time.

This critical understanding could be obtained by a survey including as many people as possible; however, such a study would be both time-consuming and expensive. Therefore, we turned to infodemiology [5], which has been employed for a long time to monitor public health issues and people's status [6,7]. During the current COVID-19 pandemic, numerous infodemiology studies have deepened our understanding of human behavior related to COVID-19 [8]. The social media platform Twitter is widely applied in the field of infodemiology [9,10]. In the field of plastic surgery, Twitter has been deployed to investigate public perception toward plastic surgery [11,12] and the engagement of plastic surgeons with social media [13,14]. Therefore, Twitter data can be used to examine public thoughts in a rapid and economic way. Thus, we collected tweets related to plastic procedures, and these data may reveal public interest in these procedures to some extent.

In this study, we extracted 73,963 publicly accessible tweets about the most common plastic procedures from January 1 to July 22, 2020. Tweeting frequency, sentiment, and word frequency analyses were applied to examine the changes in public interest in plastic procedures. This study may help stakeholders in the field of plastic surgery refocus their practices and sustain operations during the pandemic.

Methods

Data Acquisition

Tweets posted from January 1 to July 22, 2020 UTC were retrieved from Twitter using Locoyposter [15]. Locoyposter is a commercial web scraping tool that provides a visual interface that is friendly to users with little programming experience. It provides a framework in which users can collect data, such as posts on Twitter, by designing a workflow, parsing the target webpages, specifying XPath expressions, and storing the content in an external database. Keywords were determined by referring to the top plastic surgery procedures in the latest annual Plastic Surgery Statistics published by the American Society of Plastic Surgeons [16]. The top five cosmetic surgical procedures and the top five minimally invasive procedures were extracted. Among these procedures, the academic term *botulinum toxin type A* did not appear to be well known among the public, as it only returned a few tweets; therefore, the search term *Botox* was adopted because the commercial preparation Botox comprised the largest portion of botulinum toxin type A treatments according to the statistical report. The term *soft tissue fillers* was substituted with *hyaluronic acid* for the same reason.

This study focuses on firsthand and self-revealed interest in plastic procedures. Therefore, we further filtered the collected tweets by excluding replies and tweets with links. For tweets including links, interpretation should be made by referring to former tweets based on the specific conversation context or external webpages. These tweets are heterogeneous compared to their counterparts, which are self-explanatory. Introduction of these tweets would create uncertainty in this study. Only tweets in English were retained in the downstream analysis. The final query is shown below:

> *("Breast Augmentation" OR Liposuction OR Rhinoplasty OR "Eyelid Surgery" OR "Tummy Tuck" OR "Botox" OR "Hyaluronic acid" OR "Chemical Peel" OR "Laser Hair Removal" OR Microdermabrasion) lang:en until:2020-07-22 since:2020-01-01 -filter:links -filter:replies.*

This tweet search can be reproduced by pasting this query into the search box on Twitter.

We also searched these keywords on Google Trends [17], which has previously been applied in research on COVID-19 [18,19] and cosmetic procedures [20]. We searched these keywords as topics to include as many related searches as possible. The region was set as *worldwide* and the category as *all*. The default web search was selected, and the time span was January 1 to July 22, 2020. Because Google Trends normalizes data when multiple keywords are searched together [21], we searched each keyword separately. Google Trends identified "eyelid surgery" under the topic of blepharoplasty and "tummy tuck" under the topic of abdominoplasty. Therefore, these keywords were used for searches instead of the original keywords. Google Trends failed to identify any related topics for Botox; therefore, we used *Botox* as the only search term related to botulinum toxin type A. Visualization of the Google Trends of these keywords could aid division of the tweets into groups and understanding of the corresponding Twitter data.

Tweeting Frequency

Tweeting frequency was used to indicate public interest in plastic procedures. It is clear that users may tweet more frequently about certain plastic procedures if they are interested in them. We compared the overall tweeting frequencies in different phases of the pandemic. The frequencies of specific procedures and constituent ratios for all procedures were also determined. Proportions of tweets that mentioned COVID-19 were also compared by calculating the co-occurrence rate of *covid* or *coronavirus* in the tweets.

Sentiment Analysis

Sentiment analysis was performed using the Natural Language Toolkit (NLTK) [22]. We first tokenized the tweets and removed stop words and punctuation. Then, the sentiments were determined by the VADER (Valence Aware Dictionary and Sentiment Reasoner) module, which is specifically attuned to sentiments expressed in social media [23]. We compared sentiments among different phases. We also compared sentiment differences for tweets that mentioned COVID-19 to reveal

overall sentiment changes caused by COVID-19. Tweets that mentioned COVID-19 were tagged if they contained *covid* or *coronavirus* in their text. Sentiment changes for specific procedures and overall sentiment constituent ratio changes were also determined.

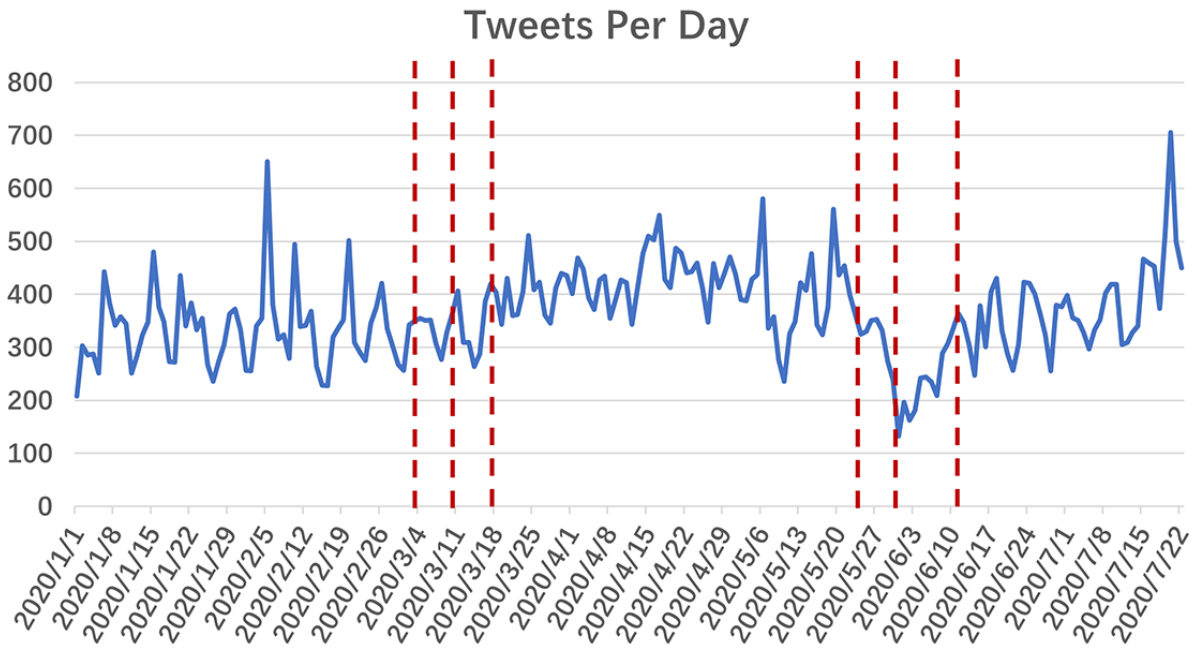
Word Frequency

Word frequency can reflect trends in topics on Twitter [24,25]. After tokenization and removal of stop words and punctuation, all words were transformed to lowercase, labeled with a part-of-speech tag, and lemmatized. Then, the word frequencies were counted and compared among different phases after normalization by the total number of tweets. A change in the frequently used words may reflect a change in public interest. Therefore, words with the highest change ratios were highlighted. These steps were achieved with the Python [26] modules NLTK and WordCloud. To better understand the word frequency results, we also provided some typical tweets for readers.

Grouping and Statistical Analysis

The collected tweets were divided into three phases. After scrutinizing the data, we found that the tweeting frequency and Google Trends indices fluctuated drastically around two significant events. The first event is the declaration of COVID-19 as a pandemic by the World Health Organization (WHO) on March 11, 2020. After this event, the amount of public attention being paid to COVID-19 increased substantially. The second significant event is the death of George Floyd [27]. The public outcry against this event drew attention away from the pandemic for a period of time. Therefore, we excluded data around these two events and divided the tweets into three phases: phase 1 (January 1 to March 4, 2020), phase 2 (March 18 to May 24, 2020), and phase 3 (June 11 to July 22, 2020). More details are provided in Figure 1.

Figure 1. Frequency of tweets related to the top plastic procedures. The dashed red lines indicate landmark events and the days around them that were excluded from the data collection process.



Quantitative variables were analyzed by one-way analysis of variance and qualitative variables were analyzed by chi-square tests using SPSS version 25 (IBM Corporation). Differences were considered significant at a P value of <.05.

Results

Tweeting Frequency

We retrieved 73,963 publicly accessible tweets about the most common plastic procedures from January 1 to July 22, 2020. We provide these tweets and their corresponding publication dates in Multimedia Appendix 1. Other information was not provided to protect the privacy of the tweet writers.

The tweeting frequency per day is shown in Figure 1. The figure shows that the tweeting frequency is different in each of the three phases. After the WHO declared COVID-19 a pandemic, the tweeting frequency increased (ie, the tweeting frequency in phase 2 is higher than that in phase 1). The tweeting frequency then decreased sharply after May 25, and the tweeting frequency in phase 3 is lower than that in phase 2. Some peaks can also be seen in the curve; these peaks were basically related to celebrities and to speculation that they may have undergone cosmetic procedures. For example, the peak on February 5 was related to Nancy Pelosi, and that on July 20 was related to Kamala Harris. More details are provided in Multimedia Appendix 1.

We also searched the top plastic procedures with Google Trends (Figure 2). It can be inferred that searches related to these plastic procedures decreased sharply after the WHO declared COVID-19 a pandemic and rebounded slowly after that. Public searches related to the majority of these procedures did decrease

at the end of May and the beginning of June. However, the decrease was less substantial than that of its Twitter counterpart.

Further statistical analysis was consistent with the results in Figure 1. As shown in Figure 3A, overall tweeting frequency per day in phase 2 increased by as much as 24.0% compared to phase 1 (333.17 vs 413.20), while tweeting frequency decreased by 9.1% after that (413.20 vs 375.33). When comparing the frequencies of tweets related to specific procedures in phase 2 and phase 1, most of them increased; only the tweeting frequencies for *eyelid surgery*, *laser hair removal*, and *microdermabrasion* showed no statistical difference. For phase 3 and phase 2, the frequency of tweets related to most procedures increased or remained stable, while the tweeting frequencies for *Botox*, *chemical peel*, and *microdermabrasion* decreased (Figure 3B).

In addition to the direct comparison of the tweeting frequencies, their relative changes were determined with a constituent ratio. As shown in Figure 3C, the overall constituent ratios were different in the different phases (χ^2=592.61, P<.001). We further applied partition of the chi-square test to each specific procedure with P values adjusted by Bonferroni corrections using SPSS. It can be inferred that the constituent ratios of *liposuction* and *tummy tuck* increased in both phase 2 and phase 3. The constituent ratios of *breast augmentation*, *eyelid surgery*, and *laser hair removal* increased in phase 3 versus phase 2, while the constituent ratios of *Botox* and *chemical peel* decreased in phase 3 versus phase 2. It should be noted that one tweet may mention more than one plastic procedure; therefore, the sum of the frequencies in Figure 3C is slightly higher than the overall number of tweets in each phase.

The proportion of tweets mentioning COVID-19 can reflect public concern to some extent. As shown in Figure 3D, the percentage of tweets that mentioned COVID-19 decreased from 2.1 in phase 2 to 0.9 in phase 3. In phase 1, only a few tweets mentioned COVID-19. Therefore, the corresponding data were not analyzed.

Figure 2. Google Trends indices for the 10 most common cosmetic surgical and minimally invasive plastic procedures. The searched topics are shown rather than the actual keywords.

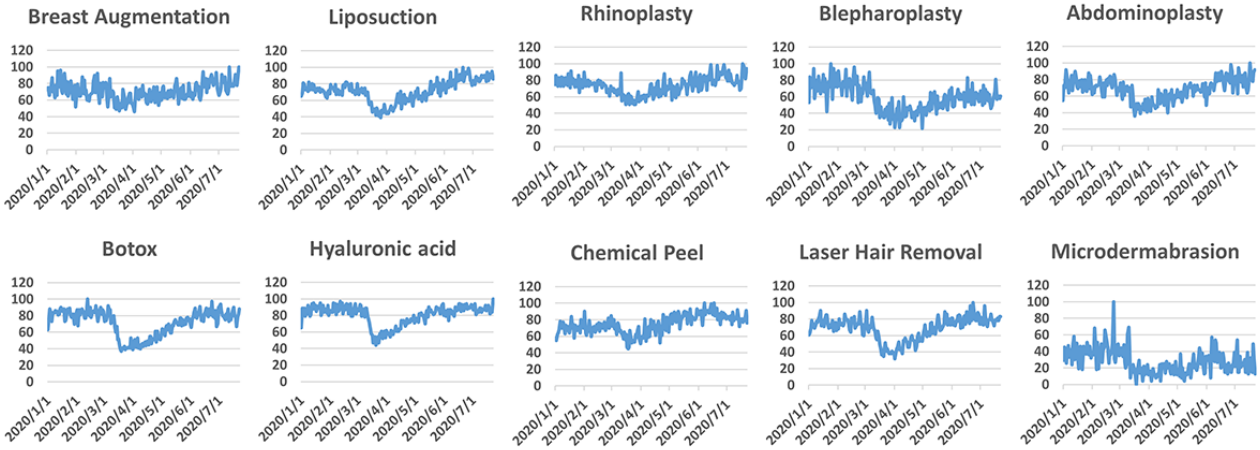
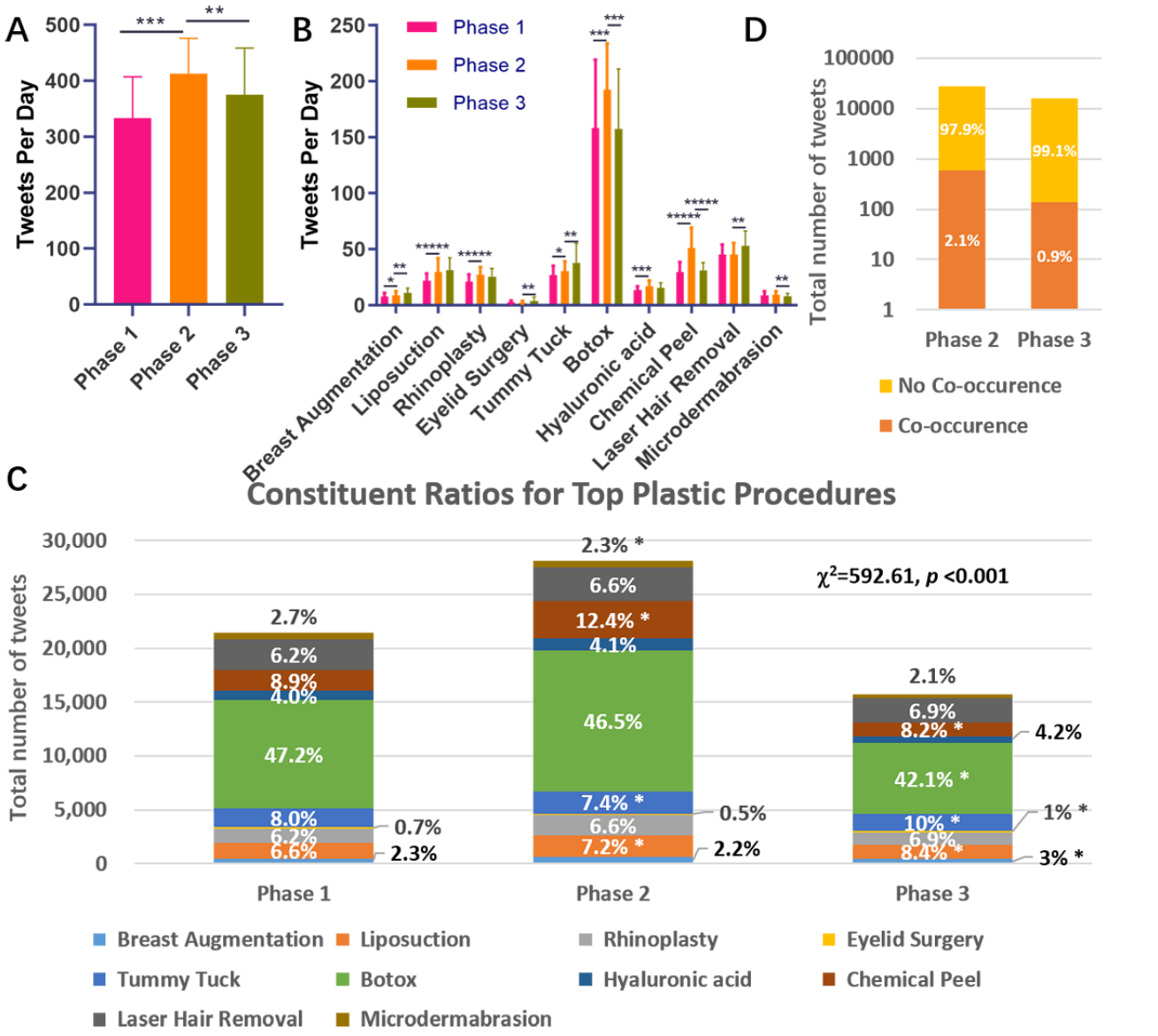


Figure 3. (A) Overall tweeting frequencies in each phase. (B) Tweeting frequencies for specific procedures. (C) Total numbers and constituent ratios (labels in cells) of tweeting frequency for the top plastic procedures. (D) Frequencies and constituent ratios (labels in cells) for tweets in which mentions of the procedures do and do not co-occur with mentions of COVID-19 in phases 2 and 3. *P<.05, **P<.01, ***P<.001.

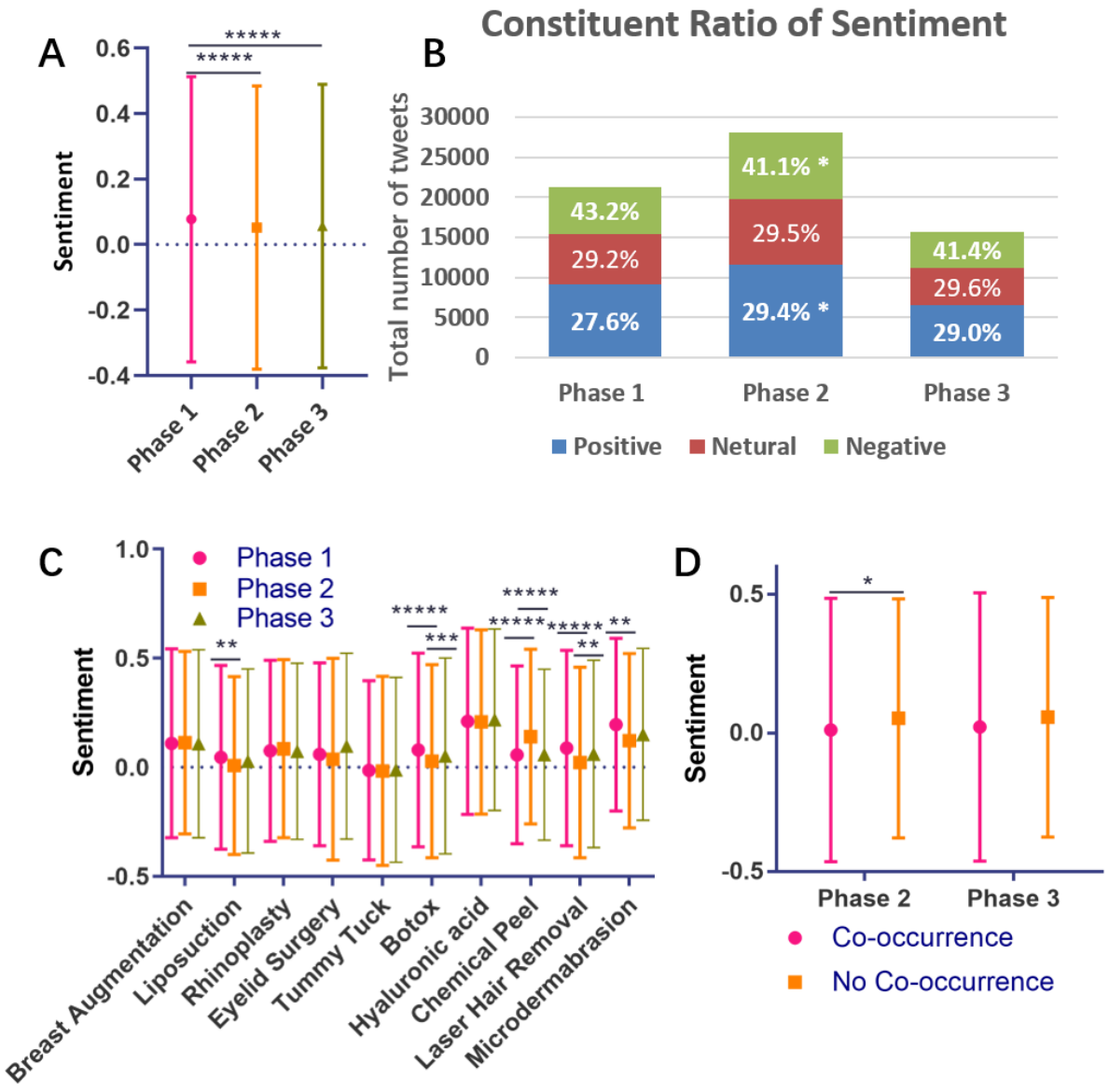


Sentiment Analysis

In addition to tweeting frequency, sentiment can serve as an indicator of public interest toward plastic procedures. As shown in Figure 4A, in general, the sentiments of tweets became more negative in phase 2 and showed no sign of rebounding in phase 3. A constituent ratio analysis found higher negative sentiment proportions and lower positive proportions in phase 2 and phase 3 than in phase 1 (Figure 4B), which is in line with the overall results. The sentiment analysis for specific procedures found that sentiments about *Botox* and *laser hair removal* decreased first and then rebounded, while sentiments about *chemical peel* showed the opposite trend (Figure 4C).

We further analyzed tweets in which *covid* or *coronavirus* co-occurred with mentions of plastic procedures, and we found that in phase 2, the sentiment of these tweets was more negative than that of tweets that did not mention COVID-19; meanwhile, this difference was not statistically significant in phase 3 (Figure 4D).

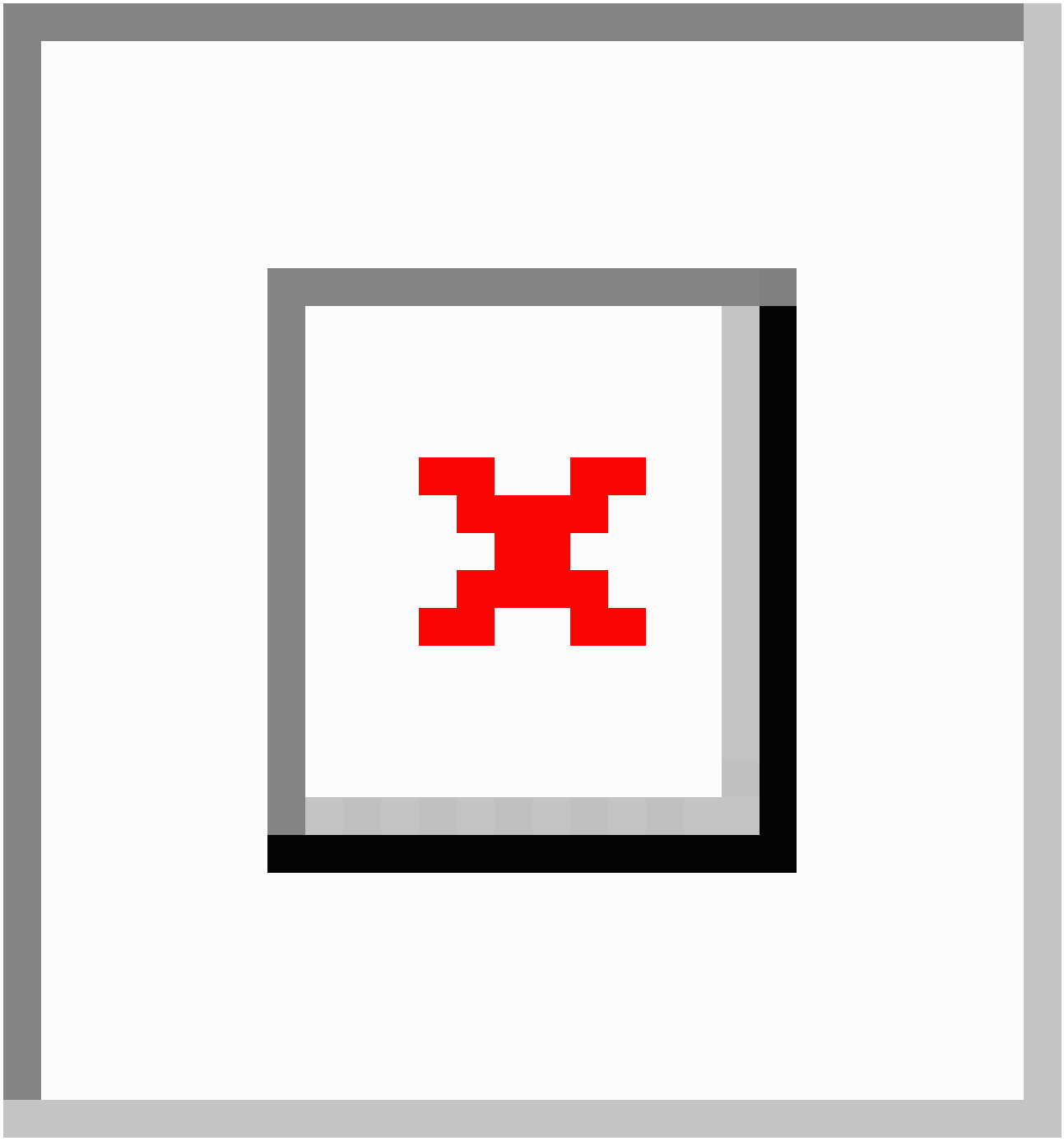
Figure 4. (A) Overall tweet sentiments in the three different phases. A negative number indicates negative sentiment, and a positive number indicates positive sentiment. (B) Total numbers and constituent ratios (labels in cells) of sentiments in different phases. (C) Sentiments for specific procedures. (D) Sentiments for tweets in which mentions of the procedures do or do not co-occur with mentions of COVID-19 in phases 2 and 3. *P<.05, **P<.01, ***P<.001.



Word Frequency

Word frequency can reflect the trending of topics, and changes in frequently used words can reveal a shift in public interest to some extent. The words with the highest frequency changes in phase 2 versus phase 1 are shown in Figure 5. It should be noted that Figure 5 was slightly manually revised to exclude some words contaminated by Twitter bots.

Figure 5. Frequencies of words with the greatest frequency changes.



The increased frequency of the words *quarantine*, *covid*, *pandemic*, *isolation*, *covid19*, *quarantined*, and *corona* indicated that COVID-19 became a major concern among people who were interested in plastic procedures. Words including *canceled*, *spas*, and *salons* usually co-occurred in phase 2. In fact, *canceled* also usually co-occurred with mentions of plastic procedures; this is not indicated in Figure 5 because this frequency change was small. The co-occurrence of these words indicates that many appointments related to plastic procedures were canceled due to the COVID-19 pandemic. Table 1 provides some typical related tweets (tweets 1 and 2). The term *morbidly* ranked high because of the surge of the hashtag #morbidlyobese. The frequency of the terms *essential* and *business* increased because many cosmetic services were considered nonessential and were cancelled, as shown in Table 1 (tweets 3 and 4). The increased frequency of the word *economy* demonstrates the public concern about the economy.

Table 1. Typical tweets related to plastic procedures. Only the tweets and their publication dates are provided to protect the privacy of their writers. The tweet contents were minimally revised to remove some foul language.

No	Date (2020)	Tweet
1	March 20	"My mom said she didn't take this pandemic serious till they canceled her laser hair removal, now she's stocking up"
2	March 20	"I was supposed to get a chemical peel today but of course that was canceled. I just wanted to have perfect skin while quarantined but I guess that's too much to ask for."
3	April 17	"I wanted to get my lips and Botox redone but apparently med spas aren't essential"
4	April 20	"Not a day goes by in which I don't feel pain that I wasn't able to get a chemical peel before non-essential businesses closed"
5	May 11	"This is the perfect time for a chemical peel. I'll at least step out for one once salons get back running"
6	May 22	"This would have been the perfect time to heal from some Botox n filler"
7	April 27	"I HATE SHAVING! The worst part of quarantine is not being able to go to my laser hair removal procedures. I was almost done with my 6 months, now I have to start all over. Shaving hurts like hell too, I'm pissed."
8	2020/4/23	"Influencers Are Still Getting Lip Fillers and Botox During Lockdown - Even during COVID-19's stay-at-home measures, spas and clinics are offering Botox and fillers at a reduced rate to influencers — and some are taking the riskier route of doing i"
9	2020/5/2	"the med spas are having covid botox sales. Brb"

The word frequencies between phase 2 and phase 3 were also compared. Some frequently appearing words are related to celebrities, such as *Kamala Harris* and *Kellyanne Conway*. Some other words are related to the Black Lives Matter movement, such as *Floyd* and *racist*. These results are not closely related to the aim of this study; therefore, they are not shown. Further details are provided in Multimedia Appendix 2. It should be noted that words related to plastic procedures, although they showed no significant frequency changes, remained the most common; this can be observed in the word cloud figure for phase 3 in Multimedia Appendix 3.

Discussion

Principal Findings

In this study, we performed a survey of 73,963 publicly accessible tweets about the most common plastic procedures from January 1 to July 22, 2020. By integrating tweeting frequency, sentiment, and word frequency analyses, we aimed to depict changes in public interest toward these plastic procedures and help stakeholders in the field of plastic surgery to sustain their operations during the difficult time of the COVID-19 pandemic. To the best of our knowledge, this study is the first to address this issue.

In this study, the keywords to be searched on Twitter were determined by referring to the annual Plastic Surgery Statistics published by the American Society of Plastic Surgeons [16]. The top five cosmetic surgical procedures and top five minimally invasive procedures certainly do not represent all plastic procedures. However, an exhaustive survey of all plastic procedures is unfeasible. The criteria used to determine which procedures to include or exclude may be controversial. Furthermore, Twitter users may use nonstandard expressions. However, these top procedures should cover most of the daily practices of many stakeholders. For the procedures that were not included, readers can refer to our methods and determine their own findings. Additionally, readers are welcome to contact us, and we will try our best to assist them.

The definition of different groups of tweets was quite difficult, as the pandemic did not occur at one static time point and evolved rapidly. There was no gold standard or even related research we could refer to. By scrutinizing trends in tweeting frequency and Google Trends indices, we found that the data fluctuated drastically around the time points of the WHO's declaration of the COVID-19 pandemic and the death of George Floyd. Naturally, we divided the collected data into three phases based on these two landmark events. The results of the tweeting frequency, sentiment, and word frequency analyses show good discrimination among these three phases. Therefore, the definition of the three phases should be suitable. There are complicated epidemiological, economic, and political reasons why these data showed distinct features in the three phases. However, these reasons are not of concern to this study; therefore, we left this question unaddressed.

In phase 2, the tweeting frequency, tweet sentiments, and word frequency all changed significantly. The tweeting frequency increased by up to 24.0%, which was somewhat surprising at first sight. Further analysis indicated that the increase was mainly contributed by the terms *Botox* and *chemical peel* (Figure 3B). Scrutiny of the related tweets revealed that many people saw the quarantine as a perfect time to receive Botox and chemical peels, as shown in Table 1 (tweets 5 and 6). This may be due to the increased free time and decreased exposure to others due to quarantine. The tweeting frequencies of most other procedures also increased (Figure 3B). Therefore, public interest in plastic procedures generally increased in phase 2, which may represent a benefit for plastic surgery stakeholders during the pandemic. However, it should be noted that the increase in public interest does not necessarily indicate that more plastic procedures were performed during this phase.

The subsequent sentiment analysis indicated that the COVID-19 pandemic resulted in negative sentiment regarding public interest in plastic procedures. However, the negative sentiment was not always detrimental for plastic surgery stakeholders. As shown in tweet 7 in Table 1, in one of the most negative tweets in phase

2, the writer expressed a strong desire for laser hair removal. The subsequent word frequency analysis basically reflected the confusion created by the COVID-19 pandemic among the public.

Upon moving to phase 3, the changes appeared to be less substantial. The overall tweeting frequency decreased but was still higher than that in phase 1. Not surprisingly, this decrease was mainly contributed by decreases in the frequency of the terms *Botox* and *chemical peel* (Figure 3B). This result may not indicate that the public was losing interest in these procedures but may be due to the decreased demand due to the restarting of the economy. In fact, people may have undergone more Botox and chemical peel procedures in phase 3 than in phase 2, which may be supported by the rebounding search indices in Google Trends (Figure 2). The discrimination between the tweeting frequency and Google Trends indices data may lie in their different natures: people express their thoughts and sentiments on Twitter, while Google is more of a tool to which people resort when they are about to take action. Our results suggest that Botox and chemical peel procedures will be the first to revive once lockdown is lifted. Because the prediction of the second wave of COVID-19 is not simply due to paranoia [28,29], stakeholders in the plastic surgery field should be prepared in case additional lockdowns are deployed.

The number of tweets that mentioned COVID-19 decreased in phase 3 (Figure 3D), and no overall sentiment difference was found (Figure 4 A, B, and D). There is also no major word frequency difference regarding plastic procedures in phase 3 versus phase 2. It appears that the public had become accustomed to coexisting with COVID-19, and their interest in plastic procedures did not decrease.

At the micro level, the tweeting frequencies of *breast augmentation*, *tummy tuck*, and *laser hair removal* increased in phase 3, and the frequencies of the first two terms also increased in phase 2. In addition to the overall tweeting frequency, the constituent ratio should also be considered. Because the damaged economy may narrow consumers' choices, plastic procedures may be required to compete with each other for consumers' favor. The constituent ratios of *liposuction* and *tummy tuck* consecutively increased in both phases 2 and 3, while the constituent ratios of *breast augmentation*, *laser hair removal*, and *eyelid surgery* only increased in phase 3. Taken together, the breast augmentation, liposuction, and tummy tuck procedures surpassed other procedures, as they showed a better absolute or relative increase. This result is not surprising, as breast augmentation and liposuction are the most popular cosmetic surgical procedures according to the latest annual Plastic Surgery Statistics published by the American Society of Plastic Surgeons [16]. Tummy tuck may not have ranked as highly as the other two procedures; however, this procedure may benefit from the effects of the current pandemic on people's lifestyles. People may be required to stay at home due to quarantine or be unable to afford to exercise at the gym due to economic concerns. Therefore, they have a higher likelihood of becoming overweight and may seek a tummy tuck or liposuction.

In summary, plastic surgery stakeholders should consider refocusing on breast augmentation, liposuction, and tummy tuck procedures at the current stage of the pandemic. If a second wave of COVID-19 occurs, stakeholders should prepare for a temporary surge of Botox and chemical peels. However, this does not mean that other procedures are unimportant, and they are still included in the majority of all plastic procedures.

When scrutinizing our data, we found that many stakeholders tried various strategies to survive the difficult period of the COVID-19 pandemic. As shown in Table 1 (tweets 8 and 9), they made use of the influence of key opinion leaders and provided more flexible prices. These efforts are praiseworthy and should be adopted by others.

Limitations

The major limitation of this study is that the "real world" is much more complicated than reflected by tweet data, even though the study is based on as many as 73,963 tweets and the research methods are well established. All surveys based on web-based social media platforms face this problem. Furthermore, the COVID-19 pandemic is unprecedented in many aspects. Because this is the first and only study of its kind to date, we did not have many other studies to refer to. Therefore, the results are open to wiser explanation by readers. Due to the limitation of Twitter privacy settings, we could not perform more precise analyses based on age, gender, or location. Additionally, the impact of the COVID-19 pandemic could vary drastically in different districts and at different time points. Readers should apply the results of this study at their own risk.

Conclusions

The public has maintained their interest in plastic procedures during the COVID-19 pandemic. Stakeholders in the field of plastic surgery should consider refocusing on breast augmentation, liposuction, and tummy tuck procedures at the current stage of the pandemic. In case of a second wave of COVID-19, stakeholders should prepare for a temporary surge in requests for Botox and chemical peels.

Acknowledgments

This work was supported by the University Key Scientific Research Projects of Henan Province (grant number 20A320033) and Medical Key Technologies R&D Program of Henan Province (grant number LHGJ20190151).

Authors' Contributions

WL conceived and designed the study, performed the data analysis, and wrote the manuscript. ZW collected, assembled, processed, and interpreted the data. XC collected and assembled the data and modified the manuscript. RP critiqued and modified the manuscript. HZ collected and processed the data. GL conceived and designed the study and had final approval of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Original tweet data. Only tweets and publication dates are provided to protect the tweet writers' privacy.
[XLSX File (Microsoft Excel File), 4700 KB - jmir_v23i3e23970_app1.xlsx]

Multimedia Appendix 2
Word frequencies in all three phases.
[XLSX File (Microsoft Excel File), 1433 KB - jmir_v23i3e23970_app2.xlsx]

Multimedia Appendix 3
Word cloud of tweets in phase 3. The size of each word is in proportion to its frequency.
[PNG File , 3168 KB - jmir_v23i3e23970_app3.png]

References

1. Armstrong A, Jeevaratnam J, Murphy G, Pasha M, Tough A, Conway-Jones R, et al. A plastic surgery service response to COVID-19 in one of the largest teaching hospitals in Europe. J Plast Reconstr Aesthet Surg 2020 Jun;73(6):1174-1205 [FREE Full text] [doi: 10.1016/j.bjps.2020.03.027] [Medline: 32359857]
2. Wang L, Gong R, Yu S, Qian H. Social media impact on a plastic surgery clinic during shutdown due to COVID-19 in China. Facial Plast Surg Aesthet Med 2020 Jun 01;22(3):162-163. [doi: 10.1089/fpsam.2020.0173] [Medline: 32298186]
3. Shokri T, Lighthall JG. Telemedicine in the era of the COVID-19 pandemic: implications in facial plastic surgery. Facial Plast Surg Aesthet Med 2020;22(3):155-156. [doi: 10.1089/fpsam.2020.0163] [Medline: 32302225]
4. Teven CM, Rebecca A. Coronavirus and the responsibility of plastic surgeons. Plast Reconstr Surg Glob Open 2020;8(4):e2855. [doi: 10.1097/gox.0000000000002855]
5. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]
6. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
7. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. J Med Internet Res 2013 Jul 18;15(7):e147 [FREE Full text] [doi: 10.2196/jmir.2740] [Medline: 23896182]
8. E-collection: JMIR theme issue 2020/21: COVID-19 special issue. Journal of Medical Internet Research. URL: https://www.jmir.org/collection/view/923 [accessed 2021-03-12]
9. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
10. Mavragani A. Infodemiology and infoveillance: scoping review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
11. Chopan M, Sayadi L, Clark EM, Maguire K. Plastic surgery and social media. Plast Reconstr Surg 2019;143(4):1259-1265. [doi: 10.1097/prs.0000000000005445]
12. Fan KL, Black CK, Luvisa K, Song DH. What does the public think? Examining plastic surgery perceptions through the Twitterverse. Plast Reconstr Surg 2019;143(2):450e-453e. [doi: 10.1097/prs.0000000000005259]
13. Branford OA, Kamali P, Rohrich RJ, Song DH, Mallucci P, Liu DZ, et al. #PlasticSurgery. Plast Reconstr Surg 2016;138(6):1354-1365. [doi: 10.1097/prs.0000000000002814]
14. Fan KL, Economides JM, Song DH. To bot or not? Challenging the top social media influencers in #PlasticSurgery. Plast Reconstr Surg 2019 Jan;143(1):337-343 [FREE Full text] [doi: 10.1097/PRS.0000000000005103] [Medline: 30589813]
15. Locoyposter. Webpage in Chinese. URL: http://www.locoyposter.com/ [accessed 2021-03-12]
16. Plastic Surgery Statistics Report. American Society of Plastic Surgeons. URL: https://www.plasticsurgery.org/documents/News/Statistics/2019/plastic-surgery-statistics-full-report-2019.pdf [accessed 2021-03-12]
17. Google Trends. URL: https://trends.google.com/ [accessed 2021-03-12]
18. Brodeur A, Clark A, Fleche S, Powdthavee N. Assessing the impact of the coronavirus lockdown on unhappiness, loneliness, and boredom using Google Trends. ArXiv. Preprint posted online on April 25, 2020 2020 [FREE Full text]
19. Doerr S, Gambacorta L. Identifying regions at risk with Google Trends: the impact of Covid-19 on US labour markets. BIS Bulletin. 2020 Apr 21. URL: https://www.bis.org/publ/bisbull08.pdf [accessed 2021-03-12]
20. Tijerina J, Morrison S, Nolan I, Vail D, Lee G, Nazerali R. Analysis and interpretation of Google Trends data on public interest in cosmetic body procedures. Aesthet Surg J 2020 Jan 01;40(1):NP34-NP43. [doi: 10.1093/asj/sjz051] [Medline: 30810161]

21. Mavragani A, Ochoa G. Google Trends in infodemiology and infoveillance: methodology framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
22. Bird SEL, Klein E. Natural Language Processing With Python. Sebastopol, CA: O'Reilly Media Inc; 2009.
23. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109/8122
24. Shahid N, Ilyas MU, Alowibdi JS, Aljohani NR. Word cloud segmentation for simplified exploration of trending topics on Twitter. IET Softw 2017 Oct;11(5):214-220. [doi: 10.1049/iet-sen.2016.0307]
25. Ainin S, Feizollah A, Anuar NB, Abdullah NAB, Sahran MNF. Examining what people tweet in relation to halal cosmetics-related topics. Cogent Arts Humanit 2020 Feb 06;7(1):1724593. [doi: 10.1080/23311983.2020.1724593]
26. Van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace; 2009.
27. George Floyd. Wikipedia. URL: https://en.wikipedia.org/wiki/George_Floyd [accessed 2020-07-29]
28. Kaliya-Perumal A, Kharlukhi J, Omar UF. The second wave of COVID-19: time to think of strategic stockpiles. Can J Public Health 2020 Aug 23;111(4):486-487 [FREE Full text] [doi: 10.17269/s41997-020-00371-w] [Medline: 32705521]
29. Xu S, Li Y. Beware of the second wave of COVID-19. Lancet 2020 Apr;395(10233):1321-1322. [doi: 10.1016/s0140-6736(20)30845-x]

Abbreviations

NLTK: Natural Language Toolkit
VADER: Valence Aware Dictionary and Sentiment Reasoner
WHO: World Health Organization

Edited by G Eysenbach, C Basch; submitted 30.08.20; peer-reviewed by A Mavragani, R Zowalla; comments to author 04.11.20; revised version received 28.01.21; accepted 18.02.21; published 16.03.21.

Please cite as:
Liu W, Wei Z, Cheng X, Pang R, Zhang H, Li G
Public Interest in Cosmetic Surgical and Minimally Invasive Plastic Procedures During the COVID-19 Pandemic: Infodemiology Study of Twitter Data
J Med Internet Res 2021;23(3):e23970
URL: https://www.jmir.org/2021/3/e23970
doi:10.2196/23970
PMID:33608248

Original Paper

Cost-effectiveness of a Telemonitoring Program for Patients With Heart Failure During the COVID-19 Pandemic in Hong Kong: Model Development and Data Analysis

Xinchan Jiang[1], BSc, MPhil; Jiaqi Yao[1], BSc, MPhil; Joyce Hoi-Sze You[1], PharmD

School of Pharmacy, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
Joyce Hoi-Sze You, PharmD
School of Pharmacy
Faculty of Medicine
The Chinese University of Hong Kong
8th Floor, Lo Kwee-Seong Integrated Biomedical Sciences Building
Shatin, NT
Hong Kong
China (Hong Kong)
Phone: 852 39436830
Email: joyceyou@cuhk.edu.hk

Abstract

Background: The COVID-19 pandemic has caused patients to avoid seeking medical care. Provision of telemonitoring programs in addition to usual care has demonstrated improved effectiveness in managing patients with heart failure (HF).

Objective: We aimed to examine the potential clinical and health economic outcomes of a telemonitoring program for management of patients with HF during the COVID-19 pandemic from the perspective of health care providers in Hong Kong.

Methods: A Markov model was designed to compare the outcomes of a care under COVID-19 (CUC) group and a telemonitoring plus CUC group (telemonitoring group) in a hypothetical cohort of older patients with HF in Hong Kong. The model outcome measures were direct medical cost, quality-adjusted life-years (QALYs), and incremental cost-effectiveness ratio. Sensitivity analyses were performed to examine the model assumptions and the robustness of the base-case results.

Results: In the base-case analysis, the telemonitoring group showed a higher QALY gain (1.9007) at a higher cost (US $15,888) compared to the CUC group (1.8345 QALYs at US $15,603). Adopting US $48,937/QALY (1 × the gross domestic product per capita of Hong Kong) as the willingness-to-pay threshold, telemonitoring was accepted as a highly cost-effective strategy, with an incremental cost-effective ratio of US $4292/QALY. No threshold value was identified in the deterministic sensitivity analysis. In the probabilistic sensitivity analysis, telemonitoring was accepted as cost-effective in 99.22% of 10,000 Monte Carlo simulations.

Conclusions: Compared to the current outpatient care alone under the COVID-19 pandemic, the addition of telemonitoring-mediated management to the current care for patients with HF appears to be a highly cost-effective strategy from the perspective of health care providers in Hong Kong.

(J Med Internet Res 2021;23(3):e26516) doi:10.2196/26516

KEYWORDS

telemonitoring; mobile health; smartphone; heart failure; COVID-19; health care avoidance; cost-effectiveness

Introduction

Heart failure (HF) is a chronic disease affecting 38 million patients worldwide, with high in-hospital mortality (6.4%), 1-year readmission rate (24%-30%), and 1-year postdischarge mortality (20%) [1-5]. This chronic cardiac disease imposes a substantial global economic burden of US $108 billion per annum (approximated in 2012) [6], which is expected to increase considerably with the aging of the population [7]. Hong Kong is a developed city with an aging population, and the local epidemiological findings on outcomes of patients with HF were consistent with those of western countries [8,9].

The COVID-19 pandemic has imposed major burdens and barriers on the operation of health care systems worldwide. COVID-19 has not only disrupted the provision of routine

medical care but has also caused patients to delay and avoid seeking medical care [10]. COVID-19 was reported to be a factor associated with avoiding medical consultation in Hong Kong [11]. Patients with chronic conditions such as HF are therefore at risk of suboptimal care during the COVID-19 pandemic as a result of disruption or avoidance of routine medical care. The treatment outcomes of HF under current care during the COVID-19 pandemic are expected to be compromised.

Telehealth is a potential timely alternative to minimize the risk of COVID-19 transmission by reducing direct physical contact and to sustain continuous medical care to patients with HF during the COVID-19 pandemic [12]. The benefits of telemonitoring programs have been examined in clinical studies for the management of patients with HF. A meta-analysis reported that the application of telemonitoring program was associated with reduced risk of all-cause mortality and HF-related mortality [13].

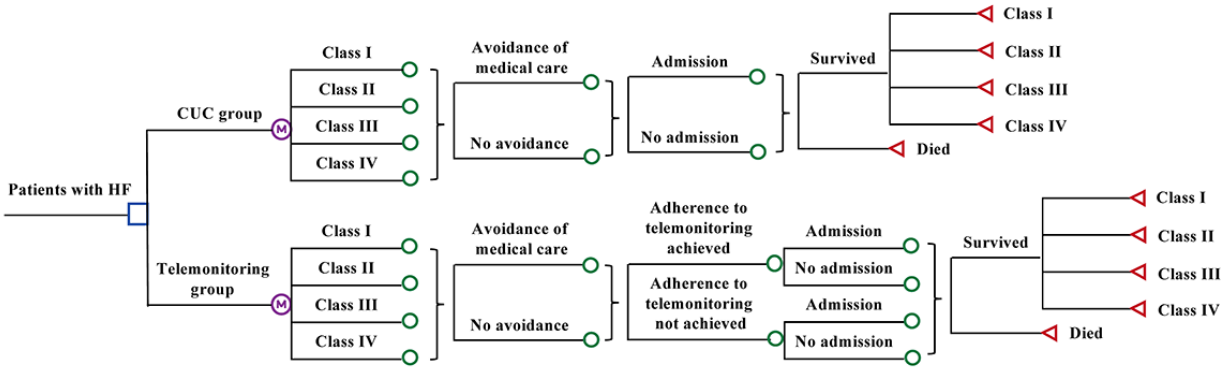
The Markov model is a well-established decision-analytic model for simulation of expected treatment costs and health-related outcomes by incorporating relevant clinical probabilities, costs, and utility inputs. In a Markov model, hypothetical subjects proceed through health states (Markov states) in the next model cycle according to transition probabilities. Markov modeling is recommended for evaluating the outcomes of diseases that might progress, improve, or relapse through transition over a series of health states [14]. The cost-effective application of telemonitoring for the management of HF was demonstrated by Markov model–based analyses prior to the era of COVID-19 [15,16], and the patients' medical avoidance was therefore not evaluated as an influential factor. In this study, COVID-related medical avoidance was considered in the model-based analysis. The aim of our study was to examine the potential clinical and health economic outcomes of adding telemonitoring programs to current medical care during the COVID-19 pandemic for the management of patients with HF from the perspective of health care providers in Hong Kong.

Methods

Model Design

A Markov decision-analytic model was designed to estimate the potential outcomes of current care under COVID-19 (CUC) with and without telemonitoring in a hypothetical cohort of older patients with HF (age 65 years or above) in Hong Kong (Figure 1). The outcomes were simulated from the entry of the model for a time frame of 10 years or until death, whichever occurred first. The two strategies examined in this study were (1) CUC plus telemonitoring (telemonitoring group) and (2) CUC alone (CUC group). The hypothetical cohort entered the model at one of the New York Heart Association (NYHA) classes I-IV and proceeded to another health status by the corresponding probability in each monthly cycle. The model outcome measures were direct medical cost, quality-adjusted life-years (QALYs), and incremental cost-effectiveness ratio (ICER).

Figure 1. Simplified Markov model of telemonitoring for patients with HF. CUC: care under COVID-19; HF: heart failure.



Multidisciplinary care is the standard management approach in usual care for patients with HF in Hong Kong, as recommended by the American College of Cardiology Foundation/American Health Association Guideline for the Management of Heart Failure [17]. Patients in the CUC and telemonitoring groups therefore all received multidisciplinary care, while patients in telemonitoring group received telemonitoring-mediated HF management in addition to multidisciplinary care. The telemonitoring-mediated management approach evaluated in a clinical outcome study was adopted in this model [18]. The patients in the telemonitoring group transmitted cardiac measures (heart rate, blood pressure, and weight) daily to the HF management team and answered a short series of questions pertinent to their HF symptoms via an app downloaded to a smartphone. A clinically validated algorithm that was embedded in the app stratified patients into different states and further identified patients with urgent needs. The patients with urgent needs would receive an alert message and an automated call suggesting emergent services. The on-call clinician would also be alerted to provide timely intervention at the onset of symptom exacerbations. Patients who were classified as nonurgent cases would receive self-instruction on administration of medications and when to contact a care provider.

Because of patients' concerns about the risk of acquiring COVID-19 at health care facilities during the pandemic, patients in both arms might or might not have avoided attending the in-person medical care clinic. The telemonitoring-mediated care also required daily transmission of cardiac measures via a smartphone app, and patients in the telemonitoring group might or might not have achieved adherence to the telemonitoring requirements. Patients in both arms might have experienced

HF-related hospitalization. For the patients who survived (with or without hospitalization) in each cycle, they might have remained in the same NYHA classification or improved/progressed to another NYHA classification.

Model Inputs

All the model inputs are shown in Table 1. The clinical inputs were retrieved from published reports written in English, identified from a literature search on MEDLINE over the period of 2000-2020. Epidemiology or disease burden studies in the Chinese population, randomized clinical trials, and meta-analyses were the preferred sources for clinical model inputs.

Table 1. Model parameters.

Parameters	Base case value	Range of sensitivity analysis	Distribution	Reference
Clinical inputs				
Proportion of NYHA[a] classification (%)			Dirichlet	[19]
Class I	9	8.1-9.9		
Class II	44	39.6-48.4		
Class III	34	30.6-37.4		
Class IV	13	8.6-17.4		
Transition probability (monthly)			Dirichlet	[20]
I to I	0.9597	0.9538-0.9678		
I to II	0.0394	0.0315-0.0473		
I to III	0.0009	0.0007-0.0011		
I to IV	0	0-0.0011		
II to I	0.0073	0.0058-0.0088		
II to II	0.9877	0.9852-0.9902		
II to III	0.0039	0.0031-0.0047		
II to IV	0.0011	0.0009-0.0013		
III to I	0.001	0.0008-0.0012		
III to II	0.0443	0.0354-0.0532		
III to III	0.8843	0.8612-0.9074		
III to IV	0.0704	0.0563-0.0845		
VI to I	0.0010	0.0008-0.0012		
VI to II	0.0443	0.0354-0.0532		
VI to III	0.8515	0.8612-0.9074		
VI to IV	0.1032	0.0563-0.0845		
Probability of HF[b]-related hospitalization in multidisciplinary care (monthly)	0.0296	0.0237-0.15	Beta	[9]
Probability of all-cause mortality in multidisciplinary care (monthly)	0.0279	0.0076-0.0383	Beta	[9]
Risk ratio of event with versus without multidisciplinary care				
HF-related hospitalization	0.74	0.64-0.87	Lognormal	[21]
All-cause mortality	0.75	0.59-0.96	Lognormal	[21]
Risk ratio of event with versus without telemonitoring				
HF-related hospitalization	0.5	0.36-0.64	Lognormal	[18]
All-cause mortality	0.81	0.70-0.94	Lognormal	[13]
Adherence to telemonitoring-guided management (%)	80	64-96	Triangular	[22]
COVID 19–related health care avoidance (%)	26.1	21-31.5	Triangular	[11]
Duration of COVID 19–related health care avoidance (years)	1.5	0.5-2	Triangular	[23]
Utility inputs				
Utilities			Uniform	[24]
NYHA class I	0.82	0.78-0.85		
NYHA class II	0.74	0.69-0.75		

Parameters	Base case value	Range of sensitivity analysis	Distribution	Reference
NYHA class III	0.64	0.55-0.77		
NYHA class IV	0.46	0.41-0.61		
Disutilities of hospitalization			Uniform	[24]
NYHA class I	0.04	0.03-0.05		
NYHA class II	0.07	0.06-0.08		
NYHA class III	0.10	0.08-0.12		
NYHA class IV	0.29	0.23-0.35		
Cost inputs				
Daily cost of hospitalization (US $)	654	523-785	Gamma	[25]
Length of hospitalization for HF (days)	8	6-10	Triangular	[26]
Monthly outpatient cost for HF (US $)	197	158-236	Gamma	[27]
Telemonitoring-mediated care (US $)				
Site implementation cost per patient	80	64-96	Gamma	[16]
Monthly cost of telemonitoring	50	40-60	Gamma	[16]

[a]NYHA: New York Heart Association.

[b]HF: heart failure.

At the entry of the model, the distribution of patients among the four statuses (NYHA class I: 9%, NYHA class II: 44%, NYHA class III: 34%, and NYHA class IV: 13%) adopted the baseline characteristics of patients with HF in Northeast Asia [19]. The yearly transition rates between NYHA classes were retrieved from the Eplerenone in Mild Patients Hospitalization And Survival Study in Heart Failure [20], and MATLAB (MathWorks) was used to generate the monthly transition matrix. HF-related hospitalization (2.96%) and all-cause mortality for patients aged ≥65 years (2.79%) with multidisciplinary care were approximated from the Hong Kong Heart Failure Registry. In this study, a total of 1940 new-onset HF cases were identified in the Hong Kong Chinese population between 2005 and 2012. Both of the above estimates were retrieved from patients followed in the outpatient setting, with a prior history of hospitalization for decompensated HF [9]. The clinical impacts of multidisciplinary care (vs without multidisciplinary care) on HF-related admission (risk ratio [RR] 0.74; 95% CI 0.63-0.87) and all-cause mortality (RR 0.75; 95% CI 0.59-0.96) were retrieved from a systematic review of 29 trials (5039 patients) on multidisciplinary strategies for management of patients with HF [21]. The probabilities of HF-related hospitalization and all-cause mortality in patients who avoided medical care during the COVID-19 pandemic were approximated using the risks of events without multidisciplinary care. The relative change of hospitalization rate associated with telemonitoring-medicated care (RR 0.5, 95% CI 0.36-0.64) was obtained from an outcome study of a smartphone-based telemonitoring system in 315 patients with HF [18]. The relative impact of telemonitoring on all-cause mortality (RR 0.81, 95% CI 0.70-0.94) was estimated from a meta-analysis of 37 trials that evaluated the comparative effectiveness of telemonitoring versus no telemonitoring for HF management [13]. The adherence of telemonitoring was defined as achieving 70% of scheduled daily data transmission and HF symptom reporting. The percentage of achieved adherence was assumed to be 80% based on a study investigating the patient adherence of a smartphone-based telemonitoring system for HF [22]. The percentage of medical avoidance among patients with HF (26.1%) was approximated from a public survey of 765 subjects on the use of health services during the COVID-19 pandemic in Hong Kong [11]. The base-case value of health care avoidance duration was estimated to be 1.5 years with a range of 0.5-2 years, based upon the epidemiologic projections of the COVID-19 pandemic [23].

Both the utility scores of the NYHA classes and disutilities due to hospitalization were retrieved from the predicted utilities of patients with HF in the Systolic Heart Failure Treatment with the I_f Inhibitor Ivabradine Trial (n=5313) [24]. The expected QALY gain in each group was calculated by the time spent in the health statuses and the corresponding utility scores. The QALY gain was discounted at an annual rate of 3%.

The cost analysis in this model was conducted using direct medical costs in the year 2020 from the perspective of public health care providers in Hong Kong. The costs of telemonitoring-medicated care (in the telemonitoring group) and the costs of HF-related inpatient and outpatient care (in both groups) were included. The cost of HF-related hospitalization was estimated by the daily cost of inpatient care and the length of stay of the patients. The daily cost of inpatient care was approximated from the fees and charges of public hospital services provided by the Hospital Authority in Hong Kong [25]. The length of hospital stay was estimated from a review on the burden of HF in 9 countries or regions (including Hong Kong) in Asia [26]. The monthly outpatient cost was estimated from the findings of a retrospective observational study on the total management cost (including hospitalization cost and ambulatory care cost) of patients with HF (n=73) recruited from a public hospital in Hong Kong [27]. The implementation cost of telemonitoring per capita (US $80) and

monthly cost of telemonitoring (US $50) were approximated from the reported costs of a smartphone-based telemonitoring system [16], including a smartphone, blood pressure monitor, weight scale, and licensing fee. The implementation cost was a one-time charge, while the monthly cost of telemonitoring was a recurrent cost for maintenance of the app. Hong Kong is a developed city with a high smartphone penetration rate of 85.5% in the overall population [28]. In this study, the monthly cost of telemonitoring was estimated at US $50 (US $1=HK $7.8), assuming the patients used their smartphones and installed the telemonitoring app. All costs were discounted annually by 3%.

Cost-effectiveness Analysis and Sensitivity Analysis

Expected costs and QALY gains were simulated for the two strategies in the base-case analysis. The ICERs were calculated using the equation (total $cost_{telemonitoring\ group}$–total $cost_{CUC\ group}$)/($QALY_{telemonitoring\ group}$–$QALY_{CUC\ group}$). As recommended by the World Health Organization in 2002, an ICER less than 1 × the gross domestic product per capita was considered to be highly cost‐effective [29]. The gross domestic product per capita of Hong Kong was US $48,937 in 2019 and was adopted as the willingness‐to‐pay (WTP) threshold [30]. A treatment alternative was preferred if (1) it was effective in saving QALYs at lower cost or (2) it was effective in saving QALYs at a higher cost with an acceptable ICER (< the WTP threshold).

Deterministic and probabilistic sensitivity analyses using Monte Carlo simulations were performed to examine the robustness of the base-case results. In the deterministic sensitivity analysis, each model input was evaluated over the range reported in the retrieved studies. If no range was specified, the parameter was examined over a range of ±20% of the base-case value. In the probabilistic analysis, 10,000 Monte Carlo simulations of each model outcome measure were generated by randomly drawing the value of all model inputs simultaneously from the distribution specified in Table 1. The probabilities of each strategy to be accepted as cost-effective in the 10,000 Monte Carlo simulations were determined against the variation of the WTP threshold (from US $0-100,000/QALY) in the acceptability curve. All analyses were performed using TreeAge Pro 2020 (TreeAge Software, Inc).

Results

Changes of Outcomes With Versus Without COVID-19–Related Health Care Avoidance

Over a time frame of 1.5 years (base-case value of health care avoidance duration), the expected direct medical cost and QALYs of the CUC group (with COVID-19–related health care avoidance) were US $7114 and 0.7960 QALYs, respectively. The expected cost and QALYs of usual care (without COVID-19–related health care avoidance) over a period of 1.5 years were US $6888 and 0.8135 QALYs, correspondingly. Compared with usual care (without COVID-19–related health care avoidance), CUC (with COVID-19–related health care avoidance) increased the cost by US $226 with a loss of 0.0175 QALYs.

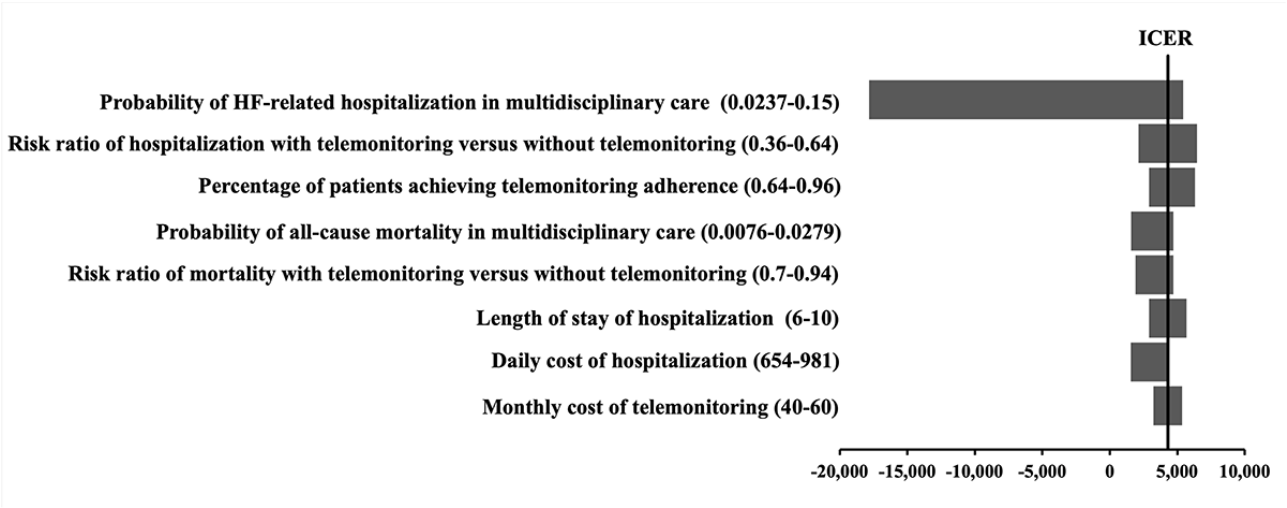
Base-Case Analysis

The expected QALY gains and total costs of the telemonitoring group and the CUC group were compared. The direct medical cost for the CUC group was US $15,603 and the QALYs were 1.8345, while these values for the telemonitoring group were US $15,888 and 1.9007, respectively. The incremental QALYs saved by the telemonitoring group (versus the CUC group) were 0.0662, with an additional cost of US $284. The ICER for the telemonitoring group versus the CUC group was US $4292/QALY, which is below the WTP threshold of 48,937 USD/QALY (1× gross domestic product per capita in Hong Kong). Telemonitoring was therefore a highly cost-effective strategy in the base-case analysis.

Sensitivity Analyses

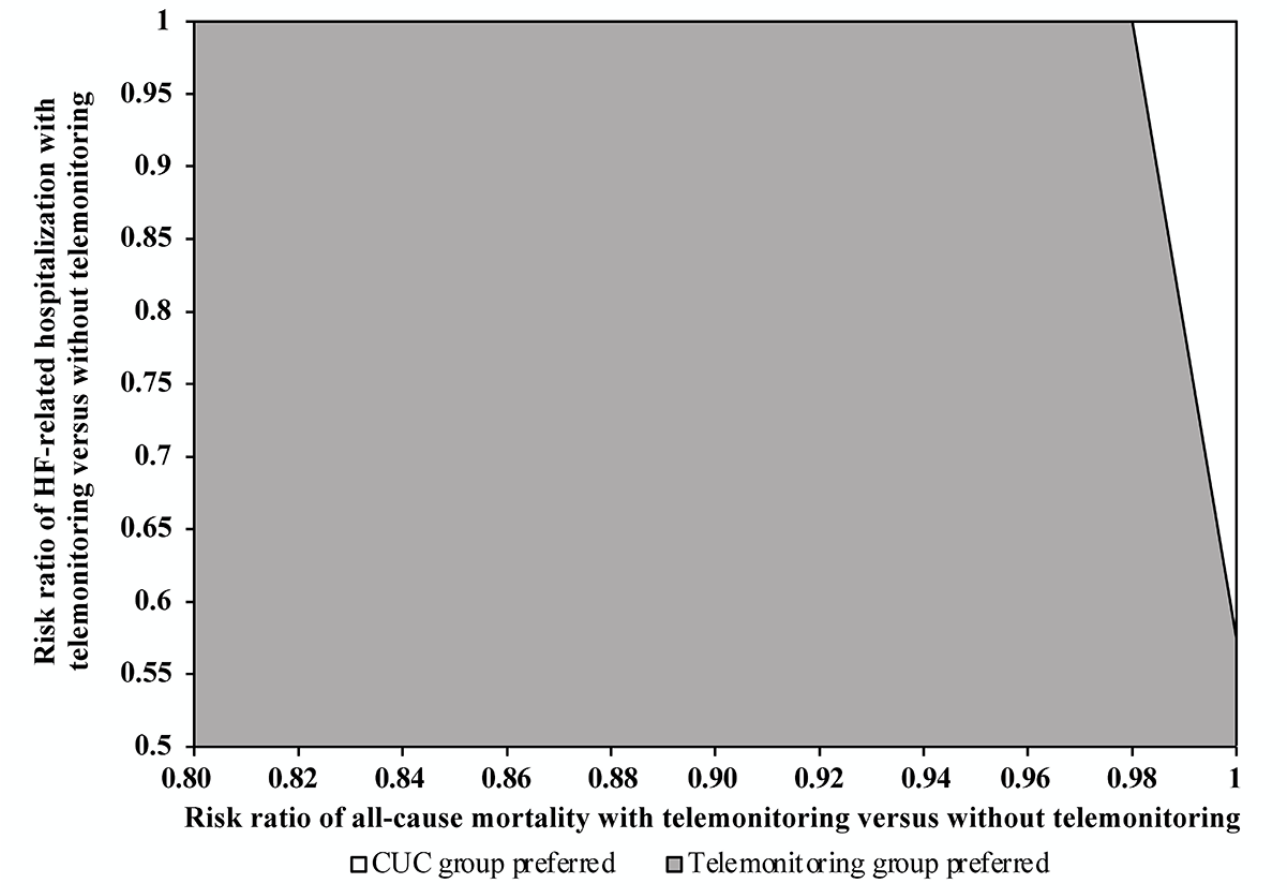
One-way deterministic sensitivity analyses were conducted for all model inputs. The ICERs of the telemonitoring group remained below the WTP threshold in the one-way variation of all parameters. No influential factor with the threshold value was found. For eight critical parameters, the ICERs varied by more than 20% (Figure 2): probability of HF-related hospitalization in multidisciplinary care, risk ratio of hospitalization with telemonitoring versus without telemonitoring, percentage of patients achieving telemonitoring adherence, probability of all-cause mortality in multidisciplinary care, risk ratio of mortality with telemonitoring versus without telemonitoring, length of stay of hospitalization, daily cost of hospitalization, and monthly cost of telemonitoring. Of these eight critical parameters, the probability of HF-related hospitalization in multidisciplinary care had the highest impact on the total cost. When the monthly probability of HF-related hospitalization in multidisciplinary care increased from the base-case value of 0.0296 to >0.0515, the telemonitoring group gained higher QALYs at a lower cost than the CUC group.

Figure 2. One-way sensitivity analysis of the ICER of the telemonitoring group versus the CUC group. CUC: care under COVID-19; ICER: incremental cost-effectiveness ratio.



The risk ratios of telemonitoring versus no telemonitoring for HF-related hospitalization and all-cause mortality were two parameters representing the relative effectiveness of the telemonitoring-mediated care. To further investigate the interaction of these two parameters with the cost-effective acceptance of telemonitoring, a two-way deterministic sensitivity analysis was conducted with the risk ratios of telemonitoring versus without telemonitoring for HF-related hospitalization (range 0.5-1) and all-cause mortality (range: 0.81-1). The gray area in Figure 3 indicates the combinations of these two variables for telemonitoring to be acceptable as the preferred option (higher QALY gained at lower cost or at higher cost with an ICER< the WTP threshold).

Figure 3. Two-way variation of the risk ratios with telemonitoring versus without telemonitoring on HF-related hospitalization and all-cause mortality. CUC: care under COVID-19; HF: heart failure.



The incremental costs versus incremental QALYs gained by telemonitoring (when compared with the CUC group) in 10,000 Monte Carlo simulations are shown in a scatter plot in Figure 4. The telemonitoring group gained an average QALY of 0.0688 (95% CI 0.0681-0.0695, *P*<.001), with a mean additional cost of US $319 (95% CI US $306-US $333, *P*<.001). In 10,000 Monte Carlo simulations, the probability of the telemonitoring group to be more effective in QALY gain and cost-saving was 23.5%. The telemonitoring group gained a higher QALY at a higher cost, with ICER<WTP (US $48,937/QALY) 75.7% of the time.

The probabilities of each strategy to be accepted as cost-effective are shown in the acceptability curve over a wide WTP range of US $0-100,000/QALY (Figure 5). The probabilities of the telemonitoring and CUC groups were the same (50%) at a WTP threshold of US $4700/QALY. The telemonitoring group was accepted to be cost-effective 99.2% of the time at the WTP threshold of US $48,937/QALY.

Figure 4. Scatter plot of the incremental cost-effectiveness ratios for the telemonitoring group versus the care under COVID-19 group. QALY: quality-adjusted life-year; WTP: willingness-to-pay.

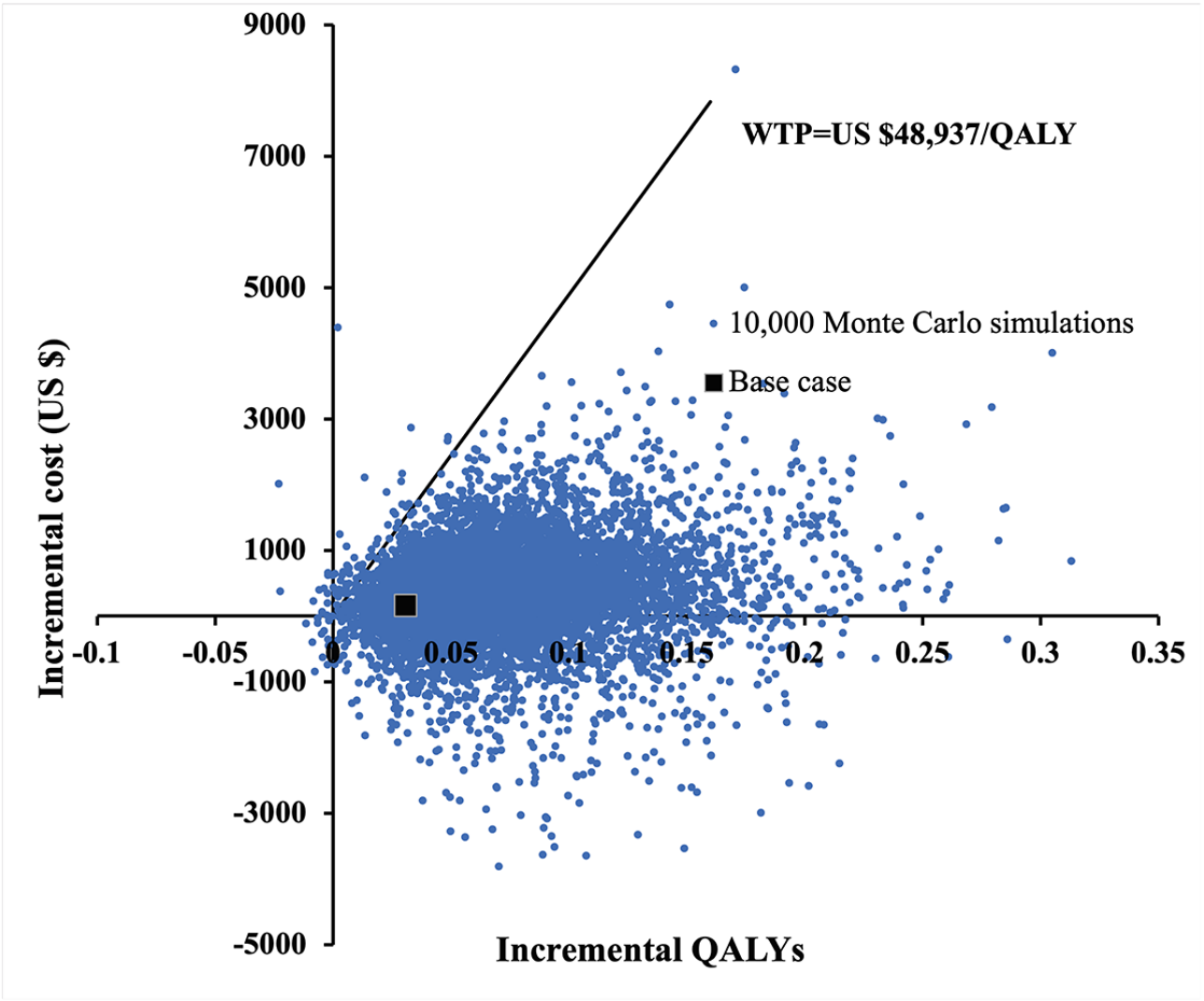
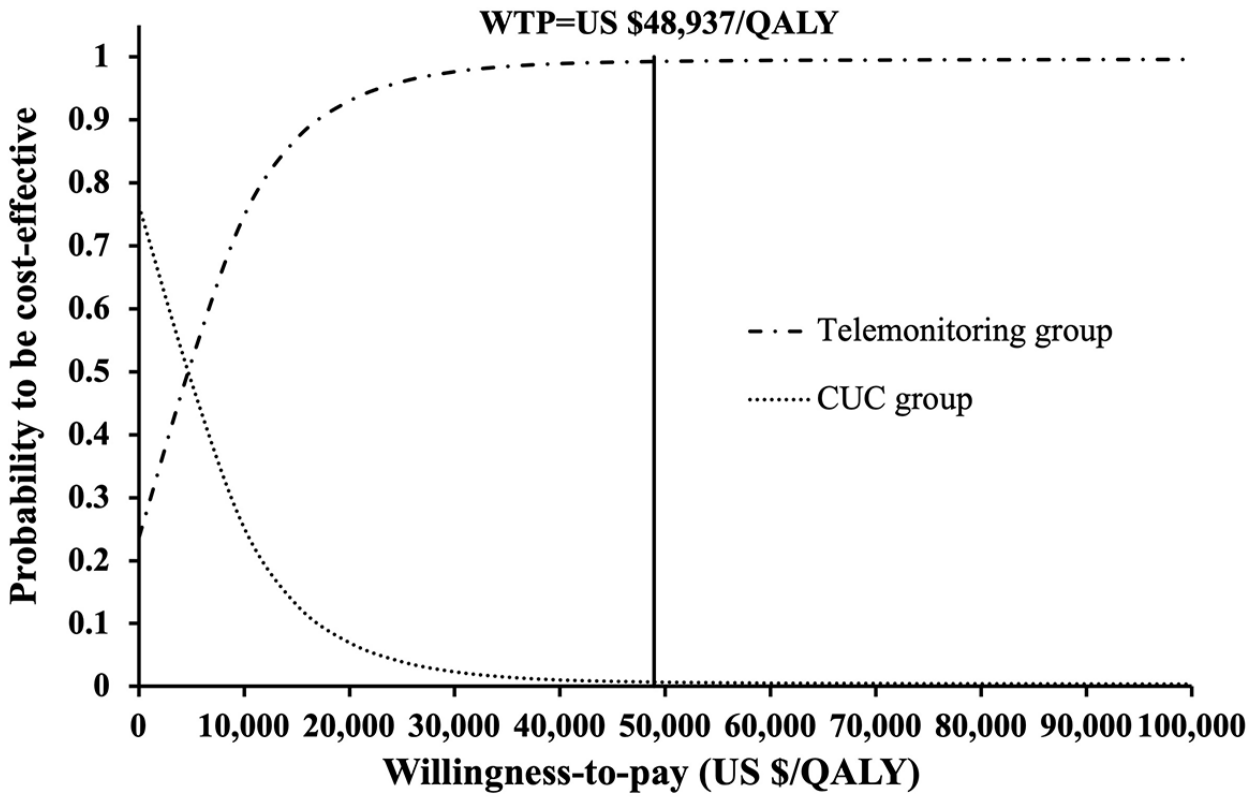


Figure 5. Cost-effectiveness acceptability curve for each strategy to be the preferred strategy against the WTP threshold. CUC: care under COVID-19; QALY: quality-adjusted life year; WTP: willingness-to-pay.



Discussion

Principal Results

This is the first analysis of the potential cost-effectiveness of smartphone-based telemonitoring systems for HF management during the COVID-19 pandemic. Our model results indicated that adding telemonitoring to current CUC for the management of patients with HF is a cost-effective strategy in the base-case analysis, with an ICER (US $4292/QALY) 10-fold below the WTP threshold (US $48,937/QALY). One-way sensitivity analysis supported the robustness of the base-case findings in that no influential parameter with a threshold value was identified. The high probability of the telemonitoring group to be accepted as the preferred strategy throughout a wide WTP range in the probabilistic sensitivity analysis further supported that adding telemonitoring to HF management is a highly cost-effective strategy.

The implementation cost is a modifiable factor when introducing a new technology in a health care system. In this study, telemonitoring was assumed to have a monthly cost of US $50 based on the estimated cost of a currently available smartphone-based telemonitoring system in Canada [16,18]. We further examined the impact of the monthly cost of the telemonitoring system in an extended one-way sensitivity analysis, and we found that telemonitoring-mediated care remained highly cost-effective if the monthly cost of telemonitoring was below US $467. Our findings were consistent with a cost-utility study of a telemonitoring-mediated HF care system in Canada in that the telemonitoring strategy was highly acceptable to be cost-effective, with an ICER of US $6701/QALY (WTP threshold=US $37,718/QALY) [16]. Our study further evaluated the interacting impact of two key parameters (risk ratios of events with telemonitoring vs without telemonitoring), which represented the relative effectiveness of telemonitoring in lowering HF-related hospitalization and all-cause mortality, on the cost-effective acceptance of the telemonitoring strategy. The combinations of these two parameters, as indicated in the two-way sensitivity analysis (Figure 3), provided the effectiveness thresholds required for the telemonitoring program to be accepted as cost-effective.

Health care systems in many countries worldwide are facing unprecedented challenges to maintaining routine medical care. This is particularly difficult when the target patients are older people with chronic cardiac diseases, who also belong to the high-risk group for life-threatening complications if they acquire COVID-19. In Hong Kong, the public health care system has struggled to provide care to patients with COVID-19 and protection against the disease to staff and other patients. Under these circumstances, public health care providers deferred some nonurgent care, and older patients also avoided attending their scheduled routine care appointments. As a result of fewer in-person clinic follow-ups, the risks of unplanned HF-related hospitalization and subsequently mortality inevitably increased.

The benefits of providing telemonitoring programs for HF management were recognized long before the COVID-19 pandemic. The pandemic has highlighted the urgency of adding telemonitoring-mediated care to in-person routine care for patients with HF [31]. Hong Kong is a developed city with a

high smartphone penetration rate [28]. An effective smartphone-based telemonitoring system with a clinician-approved algorithm is a feasible and practical option for patients with HF in Hong Kong. In light of social distancing measures in the landscape of the COVID-19 pandemic, the acceptance of applying telemonitoring-mediated care is expected to highly increase at the levels of policy decision-makers, health care providers, and patients. The COVID-19 pandemic will surely catalyze the application of telemonitoring-mediated health care services in the very near future. Cost-effectiveness evaluation of telemonitoring-based medical care is therefore highly warranted to assist policy makers in the decision-making process of resource allocation.

Limitations

There are limitations to this analysis. The cohort-based Markov model simplified real-life HF events with a limited number of health states. Other factors can impact the cost-effectiveness of HF management. For instance, influenza infection is associated with increased morbidity and mortality of patients with HF [32], and the influenza infection rate has dramatically decreased since the COVID-19 outbreak in Hong Kong [33]. Further evaluation of the impact of reduced influenza infections on HF outcome measures is highly warranted. The impact of telemonitoring on HF hospitalization and all-cause mortality varied among different types of telemonitoring, as indicated by the findings of a comprehensive network meta-analysis [13]. The cost-effectiveness of telemonitoring may therefore vary subject to the specific type of telemonitoring. Some model inputs were retrieved from overseas trials, which may affect the applicability of the model results for patients with HF in Hong Kong. Vigorous sensitivity analysis was therefore conducted on all model inputs over a broad range. The base-case results were found to be robust over the variation of all model inputs in both the deterministic and probabilistic sensitivity analyses. Additionally, the adherence of telemonitoring is not a parameter ready to be transferred between different health care systems. Health care practitioners should therefore examine the adherence of local patients when implementing a telemonitoring program for patients with HF.

Conclusion

Compared to the current CUC during the pandemic alone, the addition of telemonitoring-mediated management to current care for patients with HF appears to be a highly cost-effective strategy from the perspective of health care providers in Hong Kong. Our findings provide evidence to inform decision makers on the application of telemonitoring amid the COVID-19 pandemic. Telemonitoring has long been considered as a future model of care, and the COVID-19 pandemic has fast-forwarded the application timeline of telemonitoring in clinical settings worldwide. It is expected that a mixed mode of disease management with in-person and telemonitoring-mediated care is likely to be sustained beyond the pandemic era. Further cost-effectiveness evaluations of mixed modes of care for the management of high-burden chronic diseases, such as diabetes mellitus, are highly warranted.

Conflicts of Interest

None declared.

References

1. Braunwald E. The war against heart failure: the Lancet lecture. Lancet 2015 Feb 28;385(9970):812-824. [doi: 10.1016/S0140-6736(14)61889-4] [Medline: 25467564]
2. Lloyd-Jones DM, Larson MG, Leip EP, Beiser A, D'Agostino RB, Kannel WB, Framingham Heart Study. Lifetime risk for developing congestive heart failure: the Framingham Heart Study. Circulation 2002 Dec 10;106(24):3068-3072. [doi: 10.1161/01.cir.0000039105.49749.6f] [Medline: 12473553]
3. Nieminen MS, Brutsaert D, Dickstein K, Drexler H, Follath F, Harjola V, EuroHeart Survey Investigators, Heart Failure Association, European Society of Cardiology. EuroHeart Failure Survey II (EHFS II): a survey on hospitalized acute heart failure patients: description of population. Eur Heart J 2006 Nov;27(22):2725-2736. [doi: 10.1093/eurheartj/ehl193] [Medline: 17000631]
4. Cheng RK, Cox M, Neely ML, Heidenreich PA, Bhatt DL, Eapen ZJ, et al. Outcomes in patients with heart failure with preserved, borderline, and reduced ejection fraction in the Medicare population. Am Heart J 2014 Nov;168(5):721-730. [doi: 10.1016/j.ahj.2014.07.008] [Medline: 25440801]
5. Tromp J, Bamadhaj S, Cleland JGF, Angermann CE, Dahlstrom U, Ouwerkerk W, et al. Post-discharge prognosis of patients admitted to hospital for heart failure by world region, and national level of income and income disparity (REPORT-HF): a cohort study. Lancet Glob Health 2020 Mar;8(3):e411-e422 [FREE Full text] [doi: 10.1016/S2214-109X(20)30004-8] [Medline: 32710857]
6. Cook C, Cole G, Asaria P, Jabbour R, Francis DP. The annual global economic burden of heart failure. Int J Cardiol 2014 Feb 15;171(3):368-376. [doi: 10.1016/j.ijcard.2013.12.028] [Medline: 24398230]
7. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, American Heart Association Advocacy Coordinating Committee, Council on Arteriosclerosis, Thrombosis and Vascular Biology, Council on Cardiovascular Radiology and Intervention, Council on Clinical Cardiology, Council on Epidemiology and Prevention, Stroke Council. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circ Heart Fail 2013 May;6(3):606-619 [FREE Full text] [doi: 10.1161/HHF.0b013e318291329a] [Medline: 23616602]

8. Kelvin W, Yeung M. Population ageing trend of Hong Kong. Office of the Government Economist, The Government of the Hong Kong Special Administrative Region. 2019 Jan. URL: https://www.hkeconomy.gov.hk/en/pdf/el/el-2019-02.pdf [accessed 2020-11-24]
9. Hai J, Chan P, Huang D, Ho M, Ho C, Cheung E, et al. Clinical characteristics, management, and outcomes of hospitalized heart failure in a Chinese population-The Hong Kong Heart Failure Registry. J Card Fail 2016 Aug;22(8):600-608. [doi: 10.1016/j.cardfail.2016.03.007] [Medline: 27002944]
10. Czeisler M, Marynak K, Clarke KEN, Salah Z, Shakya I, Thierry JM, et al. Delay or avoidance of medical care because of COVID-19-related concerns - United States, June 2020. MMWR Morb Mortal Wkly Rep 2020 Sep 11;69(36):1250-1257 [FREE Full text] [doi: 10.15585/mmwr.mm6936a4] [Medline: 32915166]
11. Hung KK, Walline JH, Chan EYY, Huang Z, Lo ESK, Yeoh EK, et al. Health service utilization in Hong Kong during the COVID-19 pandemic - a cross-sectional public survey. Int J Health Policy Manag. Online ahead of print 2020 Oct 19. [doi: 10.34172/ijhpm.2020.183] [Medline: 33105965]
12. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug 01;20(1):1193 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
13. Yun JE, Park J, Park H, Lee H, Park D. Comparative effectiveness of telemonitoring versus usual care for heart failure: a systematic review and meta-analysis. J Card Fail 2018 Jan;24(1):19-28. [doi: 10.1016/j.cardfail.2017.09.006] [Medline: 28939459]
14. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices—overview. Med Decis Making 2012 Sep 18;32(5):667-677. [doi: 10.1177/0272989x12454577]
15. Thokala P, Baalbaki H, Brennan A, Pandor A, Stevens JW, Gomersall T, et al. Telemonitoring after discharge from hospital with heart failure: cost-effectiveness modelling of alternative service designs. BMJ Open 2013 Sep 18;3(9):e003250 [FREE Full text] [doi: 10.1136/bmjopen-2013-003250] [Medline: 24048626]
16. Boodoo C, Zhang Q, Ross HJ, Alba AC, Laporte A, Seto E. Evaluation of a heart failure telemonitoring program through a microsimulation model: cost-utility analysis. J Med Internet Res 2020 Oct 06;22(10):e18917 [FREE Full text] [doi: 10.2196/18917] [Medline: 33021485]
17. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Colvin MM, et al. 2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. J Card Fail 2017 Aug;23(8):628-651. [doi: 10.1016/j.cardfail.2017.04.014] [Medline: 28461259]
18. Ware P, Ross HJ, Cafazzo JA, Boodoo C, Munnery M, Seto E. Outcomes of a heart failure telemonitoring program implemented as the standard of care in an outpatient heart function clinic: pretest-posttest pragmatic study. J Med Internet Res 2020 Feb 08;22(2):e16538 [FREE Full text] [doi: 10.2196/16538] [Medline: 32027309]
19. Lam CSP, Teng TK, Tay WT, Anand I, Zhang S, Shimizu W, et al. Regional and ethnic differences among patients with heart failure in Asia: the Asian sudden cardiac death in heart failure registry. Eur Heart J 2016 Nov 01;37(41):3141-3153. [doi: 10.1093/eurheartj/ehw331] [Medline: 27502121]
20. Ademi Z, Pasupathi K, Liew D. Cost-effectiveness of eplerenone compared to usual care in patients with chronic heart failure and NYHA class II symptoms, an Australian perspective. Medicine (Baltimore) 2016 May;95(18):e3531 [FREE Full text] [doi: 10.1097/MD.0000000000003531] [Medline: 27149456]
21. McAlister FA, Stewart S, Ferrua S, McMurray JJJV. Multidisciplinary strategies for the management of heart failure patients at high risk for admission: a systematic review of randomized trials. J Am Coll Cardiol 2004 Aug 18;44(4):810-819 [FREE Full text] [doi: 10.1016/j.jacc.2004.05.055] [Medline: 15312864]
22. Ware P, Dorai M, Ross HJ, Cafazzo JA, Laporte A, Boodoo C, et al. Patient adherence to a mobile phone-based heart failure telemonitoring program: a longitudinal mixed-methods study. JMIR mHealth uHealth 2019 Feb 26;7(2):e13259 [FREE Full text] [doi: 10.2196/13259] [Medline: 30806625]
23. Moore K, Lipsitch M, Barry J, Osterholm M. Part 1: The future of the COVID-19 pandemic: lessons learned from pandemic influenza. COVID-19: The CIDRAP Viewpoint. 2020 Apr 30. URL: https://www.cidrap.umn.edu/sites/default/files/public/downloads/cidrap-covid19-viewpoint-part1_0.pdf [accessed 2020-11-24]
24. Griffiths A, Paracha N, Davies A, Branscombe N, Cowie MR, Sculpher M. Analyzing health-related quality of life data to estimate parameters for cost-effectiveness models: an example using longitudinal EQ-5D data from the SHIFT randomized controlled trial. Adv Ther 2017 Mar;34(3):753-764 [FREE Full text] [doi: 10.1007/s12325-016-0471-x] [Medline: 28205056]
25. Fees and Charges. Hong Kong Hospital Authority. URL: https://www.ha.org.hk/visitor/ha_visitor_index.asp?Content_ID=10045&Lang=ENG&Dimension=100&Parent_ID=10044&Ver=HTML [accessed 2020-11-13]
26. Reyes EB, Ha J, Firdaus I, Ghazi AM, Phrommintikul A, Sim D, et al. Heart failure across Asia: same healthcare burden but differences in organization of care. Int J Cardiol 2016 Nov 15;223:163-167 [FREE Full text] [doi: 10.1016/j.ijcard.2016.07.256] [Medline: 27541646]
27. Leung AW, Chan CY, Yan BP, Yu CM, Lam YY, Lee VW. Management of heart failure with preserved ejection fraction in a local public hospital in Hong Kong. BMC Cardiovasc Disord 2015 Feb 25;15:12 [FREE Full text] [doi: 10.1186/s12872-015-0002-8] [Medline: 25887230]

28. Usage of information technology and the internet by Hong Kong residents, 2000 to 2016. Hong Kong Census and Statistics Department. 2000. URL: https://www.statistics.gov.hk/pub/B71711FB2017XXXXB0100.pdf [accessed 2020-11-28]
29. The World Health Report 2002: reducing risks, promoting healthy life. World Health Organization. 2002. URL: https://www.who.int/whr/2002/en/whr02_en.pdf?ua=1 [accessed 2020-11-24]
30. National Income. Hong Kong Census and Statistics Department. URL: https://www.censtatd.gov.hk/hkstat/sub/sp250.jsp?tableID=030&ID=0&productType=8 [accessed 2020-11-24]
31. Abraham WT, Fiuzat M, Psotka MA, O'Connor CM. Heart failure collaboratory statement on remote monitoring and social distancing in the landscape of COVID-19. JACC Heart Fail 2020 Aug;8(8):692-694 [FREE Full text] [doi: 10.1016/j.jchf.2020.06.006] [Medline: 32731947]
32. Panhwar MS, Kalra A, Gupta T, Kolte D, Khera S, Bhatt DL, et al. Effect of influenza on outcomes in patients with heart failure. JACC Heart Fail 2019 Feb;7(2):112-117 [FREE Full text] [doi: 10.1016/j.jchf.2018.10.011] [Medline: 30611718]
33. Flu Express (volume 18), number 5 (week 5). Surveillance Division of the Communicable Disease Branch of the Hong Kong Centre for Health Protection. 2021 Feb 03. URL: https://www.chp.gov.hk/files/pdf/fluexpress_week5_04_02_2021_eng.pdf [accessed 2021-01-30]

Abbreviations

CUC: care under COVID-19
HF: heart failure
ICER: incremental cost-effective ratio
NYHA: New York Heart Association
QALY: quality-adjusted life-year
RR: risk ratio
WTP: willingness-to-pay

Edited by R Kukafka, C Basch; submitted 15.12.20; peer-reviewed by Q Zhang, CP Lau; comments to author 23.01.21; revised version received 08.02.21; accepted 19.02.21; published 03.03.21.

Please cite as:
Jiang X, Yao J, You JHS
Cost-effectiveness of a Telemonitoring Program for Patients With Heart Failure During the COVID-19 Pandemic in Hong Kong: Model Development and Data Analysis
J Med Internet Res 2021;23(3):e26516
URL: https://www.jmir.org/2021/3/e26516
doi:10.2196/26516
PMID:33656440

Original Paper

Predicting Outcomes in Patients Undergoing Pancreatectomy Using Wearable Technology and Machine Learning: Prospective Cohort Study

Heidy Cos[1*], MD; Dingwen Li[1*], BS, MS; Gregory Williams[1], BS, MA; Jeffrey Chininis[1,2], BS, MEng; Ruixuan Dai[1], BSc, MSc; Jingwen Zhang[1], BSc, MSc; Rohit Srivastava[1], BSc; Lacey Raper[1], BSc; Dominic Sanford[1,2], MD, MPH; William Hawkins[1,2], MD, FACS; Chenyang Lu[1*], PhD; Chet W Hammill[1,2*], MD, MCR, FACS

[1]Washington University in St Louis, St Louis, MO, United States

[2]Barnes-Jewish Hospital and the Alvin J Siteman Cancer Center, St Louis, MO, United States

[*]these authors contributed equally

Corresponding Author:
Chet W Hammill, MD, MCR, FACS
Barnes-Jewish Hospital and the Alvin J Siteman Cancer Center
660 S Euclid Ave
Campus Box 8109
St Louis, MO, 63110
United States
Phone: 1 3142731809
Email: hammillc@wustl.edu

Abstract

Background: Pancreatic cancer is the third leading cause of cancer-related deaths, and although pancreatectomy is currently the only curative treatment, it is associated with significant morbidity.

Objective: The objective of this study was to evaluate the utility of wearable telemonitoring technologies to predict treatment outcomes using patient activity metrics and machine learning.

Methods: In this prospective, single-center, single-cohort study, patients scheduled for pancreatectomy were provided with a wearable telemonitoring device to be worn prior to surgery. Patient clinical data were collected and all patients were evaluated using the American College of Surgeons National Surgical Quality Improvement Program surgical risk calculator (ACS-NSQIP SRC). Machine learning models were developed to predict whether patients would have a textbook outcome and compared with the ACS-NSQIP SRC using area under the receiver operating characteristic (AUROC) curves.

Results: Between February 2019 and February 2020, 48 patients completed the study. Patient activity metrics were collected over an average of 27.8 days before surgery. Patients took an average of 4162.1 (SD 4052.6) steps per day and had an average heart rate of 75.6 (SD 14.8) beats per minute. Twenty-eight (58%) patients had a textbook outcome after pancreatectomy. The group of 20 (42%) patients who did not have a textbook outcome included 14 patients with severe complications and 11 patients requiring readmission. The ACS-NSQIP SRC had an AUROC curve of 0.6333 to predict failure to achieve a textbook outcome, while our model combining patient clinical characteristics and patient activity data achieved the highest performance with an AUROC curve of 0.7875.

Conclusions: Machine learning models outperformed ACS-NSQIP SRC estimates in predicting textbook outcomes after pancreatectomy. The highest performance was observed when machine learning models incorporated patient clinical characteristics and activity metrics.

(J Med Internet Res 2021;23(3):e23595) doi:10.2196/23595

KEYWORDS

pancreatectomy; pancreatic cancer; telemonitoring; remote monitoring; machine learning; wearable technology; activity

Introduction

Pancreatectomy is a particularly complex operation with a 90-day mortality rate over 4% and serious morbidity rates over 20%, even in high-volume centers [1,2]. In the recently completed Alliance for Clinical Trials in Oncology (ALLIANCE) trial A021101 [3] and PREOPANC [4] multicenter clinical trials, 53% and 68% of patients, respectively, experienced at least a moderate complication from pancreatectomy. When a complication occurs after a pancreatectomy, the cost of the procedure to the health care system nearly triples from US $31,809 to US $82,576 because of prolonged hospitalization, additional treatments, and readmissions [5,6]. Complications are especially morbid in patients with pancreas cancer, a frail population with a mean age of 70 years, with up to 40% of patients being malnourished on presentation [7]. Multiple studies have shown that patients with pancreatic cancer who experience a therapeutic complication have decreased overall survival and quality of life [8].

Patients undergoing pancreatectomy have an increased risk of postoperative complications if they have poor preoperative physical health and overall performance [9,10]. To evaluate patients for surgery, physicians perform a physical examination in the office. This is subjective and can be misleading [11-13]. The patient's condition on that day may or may not be consistent with their general health. There are simple tests such as the 6-minute walk test or the Timed Up and Go test that can be used to determine a patient's baseline physical capacity and assess if a patient is fit for the physical demands of surgery; however, these tests have not been widely adopted [11-13]. In addition, although they are more objective than a physical examination, these tests also suffer from being a single measurement at a single time point. A more widely used surgical assessment tool is the American College of Surgeons National Surgical Quality Improvement Program surgical risk calculator (ACS-NSQIP SRC) [14-16]. It uses 20 patient-specific variables to calculate the likelihood of a patient having a complication or readmission after surgery. Although these evaluation tools are helpful, there is still a major gap in the ability to objectively measure and analyze patient health status in order to determine if the patient is fit for surgery.

Recently published data have demonstrated that telemonitoring using wearable devices with a 3-axis accelerometer and photoplethysmogram sensors can provide real-time data on patient activity metrics, which can holistically capture a patient's physical health status [17-23]. A study utilizing this technology in cohorts of patients with gastrointestinal and advanced solid malignancy undergoing chemotherapeutic treatment demonstrated an inverse association between symptom severity and patient activity, with each increase of 1000 steps per day being associated with reduced odds for severe adverse events and increased survival [24,25]. Moreover, the application of machine learning methodologies and feature engineering techniques on patient activity data have shown that human biobehavioral rhythms, semantic features, and second-order statistical features are predictors of clinical outcomes [18-23]. Prognostic models derived using machine learning methodologies in patients who underwent pancreatectomy have also been shown to perform better than traditional methods in predicting outcomes [15,16].

For patients undergoing pancreatectomy, this technology has the potential to improve patient selection. To evaluate the relationship between longitudinal patient activity bioinformatics and their effect on surgical outcomes, our team implemented a protocol in which we provided patients with wearable telemonitoring devices before undergoing pancreatectomy at our institution and evaluated predictive outcomes. Herein, we present a prospective cohort study of patients undergoing pancreatectomy over a 12-month period.

Methods

Study Population

From February 2019 to February 2020, eligible patients were recruited from multidisciplinary pancreas clinics. Both men and women and members of all races and ethnic groups were eligible for this trial. The inclusion criteria for our study included patients who (1) were scheduled to undergo pancreatic resection, (2) had access to a smartphone, (3) were at least 18 years of age, and (4) were able to understand and willing to sign an institutional review board (IRB)–approved informed consent document (IRB #201810002).

Study Design

We conducted a prospective, single-center, single-cohort trial evaluating the utility of telemonitoring devices to measure daily activity in patients undergoing pancreatectomy. The device used in this study was the Fitbit Inspire HR (Fitbit, Inc), which was selected because it provides remote data access from the device with a set frequency and enhanced granularity. It is also a waterproof, inexpensive, consumer-based device and designed to be compatible with most smartphones. At the time of consent, study patients were provided with a telemonitoring device and assisted in setting it up with their smartphone. Pancreatectomy typically took place more than two weeks after surgical consent, providing a minimum of two weeks of preoperative activity metric data. All clinical practices followed the standard of care.

Patient Activity Assessments

Our team developed software to remotely collect activity metrics from our patient telemonitoring devices that was compliant with the Health Insurance Portability and Accountability Act. This platform collected real-time patient data with 1-minute granularity. In cases of a lost connection, the wearable device saved up to 7 days of minute-to-minute activity metrics as well as accessory data (eg, battery life at last sync and time of last sync). Our informatics system performed daily audits and ran a weekly summary routine to provide the study team with the previous week's data, including yield. Yield was tracked using the total number of heart rate data points obtained during the day as a proxy for the percentage of the day the patient was wearing the device properly.

Patient Clinical Assessments

Patient clinical characteristics were collected, including demographics, comorbidities, and clinical presentation.

ACS-NSQIP SRC risk calculations were evaluated and documented.

Study Outcome Measurements

All outcome measurements were prospectively collected by the study team and recorded in the patient's secure study record. All postoperative complications were coded and graded using the Modified Accordion Grading System (MAGS) [26]. The MAGS grades complications on a scale of 1 to 6, with grade 3=severe, 4=single organ system failure, 5=multiorgan system failure, and 6=death (grades 1 and 2 complications are considered nonsevere). To ensure rigor and reproducibility, surgical complications were presented and verified at a multidisciplinary pancreas conference held every week. All postoperative complications and readmissions were collected for 30 days after hospital discharge. Complications data were then used to compute the primary outcome for our study—the textbook outcome for pancreatectomy [27]. Textbook outcome was defined as the absence of postoperative pancreatic fistulae, bile leak, postpancreatectomy hemorrhage, severe complications, readmission, and in-hospital mortality. We modified our definition of textbook outcome to allow for discharging distal pancreatectomy patients with a drain on or before day 4, the standard of care in our practice.

Data Analysis

Feature Engineering

To construct machine learning models based on activity metrics data, we applied feature engineering techniques to extract three types of features: statistical, semantic, and biobehavioral rhythmic features. We extracted first- and second-order statistical features from the daily step count, heart rate, and sleep time-series data [17]. The first-order statistical features used in our analysis were mean, maximum, minimum, skewness, and kurtosis. The second-order statistical features in medical data mining were co-occurrence features for which we generated energy, entropy, correlation, inertia, and local homogeneity. We then performed detrended fluctuation analysis (DFA) on the data, which evaluates long-range correlation of noisy time-series data, and used the root-mean-square deviation from the trend, namely the fluctuation, from DFA as the feature in our analysis. [17]. The semantic features collected provided summaries of the patient's daily activity level and sleep quality. Examples of the semantic features were time in bed, minutes to fall asleep, daily sedentary time, and daily sedentary bout count. Using the previously defined methodology, we derived and calculated biobehavioral rhythm–related features from the step count and heart rate time series [18,19]. The biobehavioral rhythmic features used in our models included stability, variability, mean of the 5 least active hours each day (L5), mean of the 10 most active hours each day (M10), amplitude (M10-L5), relative amplitude ([M10-L5]/[M10+L5]) and amplitude, phase, and midline estimating statistic of rhythm (MESOR) [20,21]. Patient clinical characteristics are potentially complementary to patient activity metrics, and we incorporated that data into the predictive models. For these categorical variables, we applied standard one-hot encoding to transfer them into features that could be used together with the features extracted from the activity metrics.

To account for variation in the study participation period (ie, time to surgery), the extracted patient activity features were unified to consistent dimensions. Biobehavioral rhythmic features were computed for the entire study participation period, and the statistical and semantic features were generated daily. In order to eliminate varying input feature dimension caused by different lengths of monitoring periods, we used mean and variance of the statistical and semantic features of a participant as the final inputs to the machine learning models.

Machine Learning Methods and Statistical Considerations

Multiple machine learning models were developed, trained, and evaluated for their ability to predict outcomes by discovering complex underlying patterns from multimodal time-series patient activity data collected from wearable devices and patient clinical characteristics. To avoid overfitting, we performed state-of-the-art "shallow" machine learning models, including random forest, gradient boosted trees (GBT), k-nearest neighbors (KNN), support vector machine (SVM) with linear kernel, and logistic regression (LR) with L1 penalty. A GBT model is an ensemble of weak decision trees that classifies the samples based on the predictions of those trees [22]. The algorithm iteratively fits a weak decision tree to the pseudo-residuals from the last iteration. We then employed regularization and feature selection to avoid overfitting and improve generalizability of the models. When implementing the GBT model, we explored established regularization techniques including controlling the complexity of the trees, applying shrinkage during the training process, and using stochastic gradient boosting. In general, an SVM model constructs an optimal hyperplane or a set of hyperplanes that can separate the samples of different classes by enforcing a large margin. It then makes predictions by deciding which side or region of the hyperplane the input sample should be on. In our implementation, we chose a linear kernel instead of other nonlinear kernels, such as a radial basis function (RBF) kernel, because the linear kernel is less likely to be overfitted in small data sets. LR with L1 penalty enforces the coefficients of less important features to be shrunk to zero, which works well for the case that has multiple features. For the feature selection in the training phase, we implemented a mixture of feature selection methods, using the chi-square statistic as the heuristic for categorical features and the *F* statistic from analysis of variance (ANOVA) for continuous features. When training the models, the hyperparameters were tuned using grid search. For example, for SVM the kernel choice and regularization strength were tuned, for GBT the coefficients of L1 and L2 regularization terms and the learning rate were tuned, and for LR the coefficients of elastic net regularization were tuned.

Leave-one-subject-out cross-validation (LOSO CV) was used for calculating the performance metrics, such as area under the receiver operating characteristic (AUROC), sensitivity, specificity, precision, and F1 score. LOSO CV was able to evaluate the model's performance on unseen patients, namely the out-of-sample accuracy [23]. Model explanation techniques were explored to study the relation between input features and predicted outcomes. We used the SHapley Additive exPlanations (SHAP) technique [28], which associates each feature with an

importance score—the Shapley value. SHAP is an established model-agnostic explanation approach that can be used to explore models from any kind of machine learning [29].

Missing Data

There were three possible causes of missing data: (1) improper wearing of the device, (2) lack of user compliance (not wearing the device), and (3) loss of connectivity for longer than 7 days. For patients with missing data, we applied a two-level imputation method to the activity metrics collected by our telemonitoring devices [17]. The data-level imputation was to fill the missing data points in heart rate time series if the daily data yield, defined as the fraction of the expected data points that were successfully collected, was equal to or above the threshold (10%). The imputed time-series data were then used to compute the features [23]. We applied KNN imputation to estimate the missing heart rate data based on recent step count and heart rate data in a sliding window (eg, 5 minutes). For those heart rate time series with a daily yield of less than 10% but greater than 0%, we used feature-level imputation to directly impute their corresponding statistical and semantic features. For the feature-level imputation, we again applied KNN imputation to the missing statistical and semantic features based on other available features from the same participant on the same day. Days with no data (daily yield of 0%) were discarded in the analysis.

Model Performance Evaluation

To evaluate the effectiveness of the machine learning models in predicting postoperative outcomes, defined by the modified textbook outcome, we compared them with clinical patient performance status assessment tools, including the ACS-NSQIP SRC. Utilizing the ACS-NSQIP SRC as our baseline model, we evaluated the performance and efficacy of this approach and applied machine learning models to (1) patient clinical characteristics (demographics, comorbidities, and clinical presentation), (2) features derived from remotely collected activity metrics, and (3) patient clinical characteristics + features derived from remotely collected activity metrics. The comparative evaluation of the "patient activity–only" and "clinical characteristic–only" models assessed the predictive power of activity metrics, while the performance of a combined "patient activity + clinical characteristic" model, by design, tested whether activity metrics and clinical records complement each other to yield better results.

Results

A total of 54 patients were enrolled in the study, and 48 patients completed it. Four patients had their pancreatectomy cancelled on the day of surgery because of intraoperative evidence of advanced disease, and 2 patients electively chose to withdraw for nonmedical reasons. All patients had an independent functional status. Of the 48 patients who completed the study, 29 (60%) were females and 19 (40%) were males, with an average age of 63.2 (SD 11.6) years. Patients underwent three different types of pancreatectomy, including pancreaticoduodenectomy (n=41, 85%), distal pancreatectomy (n=6, 13%), and total pancreatectomy (n=1, 2%). The surgeries were performed open in 28 (58%) cases and minimally invasively in 20 (42%) cases. Final surgical pathology was adenocarcinoma (n=36, 75%), neuroendocrine (n=7, 15%), benign disease (n=4, 8%), and metastatic renal cell carcinoma (n=1, 2%).

In our cohort, 28 (58%) patients had a textbook outcome, with the other 20 (42%) patients not achieving a textbook outcome. Fourteen patients developed 19 severe complications (MAGS score ≥3), including delayed gastric emptying (n=3), pancreatic fistula (n=3), organ space infection (n=2), postpancreatectomy hemorrhage (n=4), nonpancreatic anastomotic leak (n=1), myocardial infarction (n=1), and other (n=5). Additionally, 11 patients required readmission to the hospital. See Table 1 for univariate analyses of demographic and comorbidity features stratified by textbook outcome in our cohort.

Table 1. Patient characteristics.

Characteristic	Patients with complications (n=20)	Patients with textbook outcomes (n=28)	*P* value[a]
Age (years), mean (range)	67.24 (48.14-80.52)	60.26 (31.02-84.02)	.04
Gender, n (%)			.12
Male	11 (55)	8 (29)	
Female	9 (45)	20 (71)	
Race, n (%)			.86
White	19 (95)	25 (89)	
Non-White	1 (5)	3 (11)	
Comorbidities, n (%)			.06
≥5	12 (60)	8 (29)	
<5	8 (40)	20 (71)	
Tobacco use, n (%)			.45
Never smoked	11 (55)	19 (68)	
Active smoker with >10 pack years	1 (5)	3 (11)	
Active smoker with <10 pack years	0	1 (3.5)	
Past history of smoking with >30 pack years	7 (35)	4 (14)	
Past history of smoking with <30 pack years	1 (5)	1 (3.5)	
Medications, n (%)			.48
≥5	7 (35)	6 (21)	
<5	13 (65)	22 (79)	
ASA[b] class, n (%)			.07
1	0	1 (3.6)	
2	7 (35)	18 (64.3)	
3	13 (65)	9 (32.1)	
BMI (kg/m^2), mean (range)	27.99 (20.30-37.00)	29.03 (19.00-48.07)	.59
Prior surgery, n (%)			.02
Yes	15 (75)	10 (36)	
No	5 (25)	18 (64)	
Operative approach, n (%)			.38
Open	14 (70)	14 (50)	
Laparoscopic	4 (20)	9 (32)	
Robotic	2 (10)	5 (18)	
Operation type, n (%)			.22
Pancreaticoduodenectomy	18 (90)	23 (82)	
Distal pancreatectomy	1 (5)	5 (18)	
Total pancreatectomy	1 (5)	0 (0)	

[a]*P* values were derived from chi-square tests for categorical variables and *F* tests for continuous variables.

[b]ASA: American Society of Anesthesiologists.

Patient activity metrics were collected over an average of 25.9 days (range 6 to 153 days) before surgery. The average daily yield of all patients, defined as the fraction of expected heart rate readings per minute that were successfully collected in a day, was 82.1% (SD 23.5%). High data availability was defined as days with a yield greater than or equal to 50%. Based on this, the average number of days per patient with high data availability was 19 (range 2 to 102) and the average percentage of days with high data availability per patient was 79.8% (range 14.8% to 100%). Patients took on average of 4162.1 (SD 4052.6)

steps per day, had an average heart rate of 75.6 (SD 14.8) beats per minute, and had an average sleep time series of 2 (SD 1), which was a mean DFA of their sleep stages with 50-minute windows. The average ACS-NSQIP SRC calculations for a patient developing any complication was 27.3% (SD 6.4%), developing a serious complication was 23.3% (SD 5.5%), and being readmitted was 15.1% (SD 3.4%).

Utilizing the ACS-NSQIP SRC as our baseline model, we evaluated the performance and efficacy of this approach and applied machine learning models to (1) patient clinical characteristics, which included demographics, comorbidities, and clinical presentation; (2) patient activity with features derived from remotely collected activity metrics; and (3) patient clinical characteristics + patient activity with features obtained or derived from both clinical records and activity metrics. Table 2 shows the performance comparison of these models at predicting a textbook outcome. The predictive models were trained with probabilistic outputs and then the classification thresholds were adjusted to obtain a sensitivity of 0.9 in order to ensure a high detection rate and allow an equitable comparison. Our AUROC curves were 0.6333 for the ACS-NSQIP SRC, 0.7054 for the patient clinical characteristics model, 0.7027 for the patient activity model, and 0.7875 for the patient clinical characteristics + patient activity model.

Table 2. Performance comparison of machine learning models trained with different data sources.

		Metrics[b]				
Parameter[a]	Model	AUROC[c] curve	Sensitivity	Specificity	Precision	F1 score
ACS-NSQIP SRC[d]		0.6333	0.9000	0.0370	0.4091	0.5625
Patient clinical characteristics	LR[e]	0.7054	0.9000	0.2321	0.4558	0.6051
Patient activity	SVM[f]	0.7027	0.9000	0.2107	0.4491	0.5992
Patient clinical characteristics + patient activity	GBT[g]	0.7875	0.9000	0.3929	0.5143	0.6545

[a]Parameters used for the models are summarized in Multimedia Appendix 1.

[b]The metrics for the machine learning models represent the average across all leave-one-subject-out cross-validation folds.

[c]AUROC: area under the receiver operating characteristic.

[d]American College of Surgeons National Surgical Quality Improvement Program surgical risk calculator (ACS-NSQIP SRC) was used as the baseline model for complications from pancreatoduodenectomy.

[e]LR: logistic regression.

[f]SVM: support vector machine.

[g]GBT: gradient boosted trees.

In our analysis, we observed that 15 out of 20 features with the highest impact discovered by SHAP were from the best performing GBT model trained on patient clinical characteristics + patient activity (see Table 3 for feature exemplars).

Finally, to determine if the amount of missing data affected the performance of the classification model, the average number of days with high data availability (again, defined as days with a yield greater than or equal to 50%) for correctly classified patients was compared with that for incorrectly classified patients. The difference in the average number of days with high data availability between correctly classified patients and incorrectly classified patients was statistically insignificant (17 days, SD 10 days, versus 25 days, SD 25 days, respectively; P=0.12). This suggests that the amount of missing data did not affect the performance of the classification model.

Table 3. Analysis of variance test statistics on the features extracted from Fitbit Inspire HR (Fitbit, Inc) data.

Features[a]	Patients with complications, mean (SD)	Patients with textbook outcomes, mean (SD)	F_{46}	*P* value	SHAP[b] value
Heart rate features					
Variance of local homogeneity	6744.5286 (5055.2469)	13362.2921 (7545.2961)	11.1603	.002	1.2694
Mean of correlation	31.9993 (0.0007)	31.9996 (0.0004)	2.5324	.12	0.2338
Mean DFA[c] of heart rate with 40-minute window	22.7418 (5.3550)	24.8816 (5.0493)	1.9086	.17	0.2214
Mean of energy	202.1648 (192.6207)	140.9836 (71.2032)	2.2724	.14	0.2064
Mean of skewness	1.3182 (0.4978)	1.1065 (0.4253)	2.4006	.13	0.1787
Cosinor amplitude	6.2318 (3.3540)	7.3569 (3.6230)	1.1464	.29	0.1507
Variance of correlation	3.3737e–7 (8.8545e–7)	9.9500e–7 (2.0791e–7)	1.7977	.19	0.1500
Log Cosinor amplitude	2.2616 (0.6844)	2.4344 (0.6922)	0.7041	.41	0.1119
Mean of kurtosis	6.2530 (2.2063)	5.6795 (2.4526)	0.6640	.42	0.0558
Variance DFA of heart rate with 30-minute window	12.1549 (7.9180)	17.6321 (11.8530)	3.1035	.08	0.0476
Step features					
Variance of daily sedentary bout	0.4669 (0.2638)	0.5574 (0.3587)	0.8798	.35	0.2174
Mean of intradaily stability	0.1100 (0.0808)	0.0689 (0.0368)	5.3752	.02	0.0930
Relative amplitude	0.2948 (0.1653)	0.2097 (0.0878)	5.0969	.03	0.0662
Intradaily stability with 60-minute window	0.1341 (0.1034)	0.0788 (0.0559)	5.4469	.02	0.0428
Sleep features					
Mean DFA of sleep stages with 50-minute window	2.8834 (0.3767)	2.9634 (0.2589)	0.7294	.40	0.0471
Categorical features					
Neutrophils	50.8000 (27.5481)	31.5393 (30.4855)	4.8323	.03	0.9024
Prior surgery	0.7500 (0.4330)	0.3571 (0.4792)	8.1374	.007	0.3428
Calcium	9.2450 (0.4955)	9.6071 (0.6464)	4.2378	.05	0.2932
ASA[d] class	2.6500 (0.4770)	2.2857 (0.5249)	5.8069	.02	0.1522
Hyperlipidemia	0.6000 (0.4899)	0.3571 (0.4792)	2.8189	.10	0.0419

[a]Statistically significant features (*P* value <.05) are listed.

[b]SHAP: SHapley Additive exPlanations.

[c]DFA: detrended fluctuation analysis.

[d]ASA: American Society of Anesthesiologists.

Discussion

Principal Results

Preoperative clinical evaluation and assessment for surgical candidacy plays an essential role in postoperative outcomes. Patients who are more physically fit for surgery are less likely to experience complications. To better predict which patients will have poor outcomes, several tools have been developed and implemented over the years, including physical examination, patient demographics, laboratory values, and risk calculators; however, none of these are perfect. In this study, we used wearable telemonitoring technology in conjunction with machine learning to evaluate patient activity preoperatively and assess its ability to predict surgical outcomes.

Our models included patient clinical characteristics, patient activity, and patient clinical characteristics combined with patient activity, which we then compared with predictions from the ACS-NSQIP SRC. We found that all three of our machine learning models outperformed the baseline estimations from the ACS-NSQIP SRC. As shown in the results section, the ACS-NSQIP SRC had an AUROC curve of 0.6333 for predicting a textbook outcome after pancreatectomy, which is consistent with previous reported findings of AUROC curves in national samples [30]. Machine learning models created using the same patient clinical characteristics utilized by the

ACS-NSQIP SRC outperformed the ACS-NSQIP SRC, with an AUROC curve of 0.7054 for LR. This was similar to machine learning models that utilized only patient activity data collected from telemonitoring (AUROC curve of 0.7027 for SVM). The best results were achieved with machine learning models that combined patient clinical characteristics with patient activity data (AUROC curve of 0.7875 for GBT). This confirmed our hypothesis that machine learning technology can outperform the standard ACS-NSQIP SRC in predicting textbook outcomes in patients who had a pancreatectomy. In addition, patient activity metrics significantly improved the predictive power.

Within the machine learning model, we utilized SHAP scores to identify features with the greatest impact. Specifically, within heart rate features, the "variance of local homogeneity" in heart rate was significantly correlated with higher SHAP values. This suggests that particular attention should be paid to patients' physiological status prior to surgery. Additionally, the "mean of intradaily stability" and "relative amplitude" of steps taken [18], which pertain to the subjects' physical mobility, were also significantly associated with higher SHAP values. The definition and derivation of these features was described by Mao et al [29]. Similar to the findings of previous studies [18,21-23], incorporating patient activity data with patient clinical data increased the performance of our machine learning models. The patient clinical data that specifically improved the models' performance included neutrophil levels, calcium levels, and a history of prior surgery. The Rotterdam Study [31] found that patients with an elevated neutrophil count in relation to lymphocyte count (neutrophil to lymphocyte ratio) were independently associated with increased morbidity and mortality. Likewise, multiple authors have also shown age-related changes in calcium metabolism and found that variations in absorption of vitamin D, as well as a decreased intake of calcium, are commonly seen in the elderly [32]; 26 (54%) of the patients in this study were aged ≥65 years at the time of surgery.

Physical activity is a targetable and modifiable behavior that has been shown to improve outcomes of cancer patients undergoing chemoradiation [33-35]. Similarly, a meta-analysis of 15 randomized controlled trials with more than 400 patients showed that prehabilitation prior to major abdominal surgery led to a significant reduction in overall and pulmonary morbidity [33].

Based on our early results, we think that the combination of patient activity metrics collected preoperatively using wearable devices and machine learning models has the potential to reliably predict operative risks. In addition, by objectively tracking activity metrics and identifying areas of weakness, the data will provide targets for preoperative optimization and allow surgeons to more efficiently engage patients in their surgical care even before they undergo a major procedure. The ultimate goal is to decrease the likelihood of postoperative complications, which we believe will have a particularly large impact on patients with pancreatic cancer, a growing population with a high proportion of elderly and frail patients.

Limitations

The study was limited by a small sample size, which could potentially increase the risk of overfitting. However, as discussed in the methods section, multiple precautions were taken to reduce the effect of overfitting. We also acknowledge the risk for selection bias, as we recruited patients with access to a smartphone, which has the potential to exclude elderly patients and patients from lower socioeconomic groups.

Conclusion

Machine learning models based on preliminary data outperform standard ACS-NSQIP SRC estimates when used to predict a textbook outcome after pancreatectomy. The highest performance at this task was observed when machine learning models incorporated patient clinical characteristics and activity metrics collected with wearable telemonitoring technology. In the future, this can provide physicians with real-time actionable data that can be used to modify management of patients undergoing pancreatectomy and develop interventions to increase patient activity.

Acknowledgments

This work was supported by grants from The Foundation for Barnes Jewish Hospital and the BJC Health Systems Innovation Lab. GW is supported by the SPORE Grant 5P50 CA196510. REDCap is supported by Clinical and Translational Science Award (CTSA) Grant UL1 TR000448 and Siteman Comprehensive Cancer Center and NCI Cancer Center Support Grant P30 CA091842.

Authors' Contributions

Authors HC and DL contributed equally as co-first authors. Authors CH and CL are co-corresponding authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Parameters used for feature extraction, imputation, and models.
[DOCX File , 14 KB - jmir_v23i3e23595_app1.docx]

References

1. Hasan S, Abel S, Verma V, Schiffman S, Thakkar S, Kulkarni A, et al. Predictors of post-operative mortality following pancreatectomy: A contemporary nationwide analysis. J Clin Oncol 2019 May 20;37(15_suppl):e15706-e15706. [doi: 10.1200/jco.2019.37.15_suppl.e15706]
2. Simons JP, Shah SA, Ng SC, Whalen GF, Tseng JF. National complication rates after pancreatectomy: beyond mere mortality. J Gastrointest Surg 2009 Oct;13(10):1798-1805. [doi: 10.1007/s11605-009-0936-1] [Medline: 19506975]
3. Katz MHG, Shi Q, Ahmad SA, Herman JM, Marsh RDW, Collisson E, et al. Preoperative Modified FOLFIRINOX Treatment Followed by Capecitabine-Based Chemoradiation for Borderline Resectable Pancreatic Cancer: Alliance for Clinical Trials in Oncology Trial A021101. JAMA Surg 2016 Aug 17;151(8):e161137 [FREE Full text] [doi: 10.1001/jamasurg.2016.1137] [Medline: 27275632]
4. Versteijne E, Suker M, Groothuis K, Akkermans-Vogelaar JM, Besselink MG, Bonsing BA, et al. Preoperative Chemoradiotherapy Versus Immediate Surgery for Resectable and Borderline Resectable Pancreatic Cancer: Results of the Dutch Randomized Phase III PREOPANC Trial. J Clin Oncol 2020 Jun 01;38(16):1763-1773. [doi: 10.1200/jco.19.02274]
5. Vonlanthen R, Slankamenac K, Breitenstein S, Puhan MA, Muller MK, Hahnloser D, et al. The impact of complications on costs of major surgical procedures: a cost analysis of 1200 patients. Ann Surg 2011 Dec;254(6):907-913. [doi: 10.1097/SLA.0b013e31821d4a43] [Medline: 21562405]
6. Enestvedt CK, Diggs BS, Cassera MA, Hammill C, Hansen PD, Wolf RF. Complications nearly double the cost of care after pancreaticoduodenectomy. Am J Surg 2012 Sep;204(3):332-338. [doi: 10.1016/j.amjsurg.2011.10.019] [Medline: 22464011]
7. Gilliland T, Villafane-Ferriol N, Shah K, Shah R, Tran Cao H, Massarweh N, et al. Nutritional and Metabolic Derangements in Pancreatic Cancer and Pancreatic Resection. Nutrients 2017 Mar 07;9(3):243 [FREE Full text] [doi: 10.3390/nu9030243] [Medline: 28272344]
8. Lubrano J, Bachelier P, Paye F, Le Treut YP, Chiche L, Sa-Cunha A, et al. Severe postoperative complications decrease overall and disease free survival in pancreatic ductal adenocarcinoma after pancreaticoduodenectomy. Eur J Surg Oncol 2018 Jul;44(7):1078-1082. [doi: 10.1016/j.ejso.2018.03.024] [Medline: 29685757]
9. Wilson R, Davies S, Yates D, Redman J, Stone M. Impaired functional capacity is associated with all-cause mortality after major elective intra-abdominal surgery. Br J Anaesth 2010 Sep;105(3):297-303 [FREE Full text] [doi: 10.1093/bja/aeq128] [Medline: 20573634]
10. Snowden C, Prentis J, Jacques B, Anderson H, Manas D, Jones D, et al. Cardiorespiratory fitness predicts mortality and hospital length of stay after major elective surgery in older people. Ann Surg 2013 Jun;257(6):999-1004. [doi: 10.1097/SLA.0b013e31828dbac2] [Medline: 23665968]
11. McKenzie G, Martin RCG. Fitness Assessment and Optimization for Hepatopancreatobiliary Surgery. Optimizing Outcomes for Liver and Pancreas Surgery. In: Rocha F, Shen P, editors. Springer. USA: Springer; 2018:1-21.
12. Huisman MG, van Leeuwen BL, Ugolini G, Montroni I, Spiliotis J, Stabilini C, et al. PLoS One 2014;9(1):e86863 [FREE Full text] [doi: 10.1371/journal.pone.0086863] [Medline: 24475186]
13. Ganga HV, Jantz J. The Limitations of the 6-Minute Walk Test as a Measurement Tool in Chronic Heart Failure Patients. Rev Esp Cardiol (Engl Ed) 2016 Jun;69(6):629. [doi: 10.1016/j.rec.2016.01.024] [Medline: 27095531]
14. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. J Am Coll Surg 2013 Nov;217(5):833-42.e1 [FREE Full text] [doi: 10.1016/j.jamcollsurg.2013.07.385] [Medline: 24055383]
15. Sahara K, Paredes AZ, Tsilimigras DI, Sasaki K, Moro A, Hyer JM, et al. Machine learning predicts unpredicted deaths with high accuracy following hepatopancreatic surgery. Hepatobiliary Surg Nutr 2021 Jan;10(1):20-30 [FREE Full text] [doi: 10.21037/hbsn.2019.11.30] [Medline: 33575287]
16. Beal E, Lyon E, Kearney J, Wei L, Ethun C, Black S, et al. Evaluating the American College of Surgeons National Surgical Quality Improvement project risk calculator: results from the U.S. Extrahepatic Biliary Malignancy Consortium. HPB (Oxford) 2017 Dec;19(12):1104-1111 [FREE Full text] [doi: 10.1016/j.hpb.2017.08.009] [Medline: 28890310]
17. Li D, Vaidya J, Wang M, Bush B, Lu C, Kollef M, et al. Feasibility Study of Monitoring Deterioration of Outpatients Using Multimodal Data Collected by Wearables. ACM Trans Comput Healthcare 2020 Mar 02;1(1):1-22. [doi: 10.1145/3344256]
18. Doryab A, Dey AK, Kao G, Low C. Modeling Biobehavioral Rhythms with Passive Sensing in the Wild. In: Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2019 Mar 29 Presented at: ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies; December 2020; USA p. 1-21. [doi: 10.1145/3314395]
19. Bae S, Dey AK, Low CA. Using passively collected sedentary behavior to predict hospital readmission. In: UBICOMP. 2016 Presented at: 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2016; USA p. 616-621.
20. Mitchell JA, Quante M, Godbole S, James P, Hipp JA, Marinac CR, et al. Variation in actigraphy-estimated rest-activity patterns by demographic factors. Chronobiol Int 2017;34(8):1042-1056 [FREE Full text] [doi: 10.1080/07420528.2017.1337032] [Medline: 28650674]
21. Huang C, Madsen MT, Gögenur I. Circadian rhythms measured by actigraphy during oncological treatments: a systematic review. Biol Rhythm Res 2015 Mar 06;46(3):329-348. [doi: 10.1080/09291016.2015.1004840]

22. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. PLoS One 2018;13(7):e0201016 [FREE Full text] [doi: 10.1371/journal.pone.0201016] [Medline: 30028888]
23. Saeb S, Lonini L, Jayaraman A, Mohr DC, Kording KP. The need to approximate the use-case in clinical machine learning. Gigascience 2017 May 01;6(5):1-9 [FREE Full text] [doi: 10.1093/gigascience/gix019] [Medline: 28327985]
24. Low C, Dey AK, Ferreira D, Kamarck T, Sun W, Bae S, et al. Estimation of Symptom Severity During Chemotherapy From Passively Sensed Data: Exploratory Study. J Med Internet Res 2017 Dec 19;19(12):e420 [FREE Full text] [doi: 10.2196/jmir.9046] [Medline: 29258977]
25. Gresham G, Hendifar AE, Spiegel B, Neeman E, Tuli R, Rimel BJ, et al. Wearable activity monitors to assess performance status and predict clinical outcomes in advanced cancer patients. NPJ Digit Med 2018;1:27 [FREE Full text] [doi: 10.1038/s41746-018-0032-6] [Medline: 31304309]
26. Strasberg S, Linehan DC, Hawkins WG. The accordion severity grading system of surgical complications. Ann Surg 2009 Aug;250(2):177-186. [doi: 10.1097/SLA.0b013e3181afde41] [Medline: 19638919]
27. van Roessel S, Mackay TM, van Dieren S, van der Schelling GP, Nieuwenhuijs VB, Bosscha K, Dutch Pancreatic Cancer Group. Textbook Outcome: Nationwide Analysis of a Novel Quality Measure in Pancreatic Surgery. Ann Surg 2020 Jan;271(1):155-162. [doi: 10.1097/SLA.0000000000003451] [Medline: 31274651]
28. Lundberg SL, Lee S. A unified approach to interpreting model predictions. 2017 Presented at: Neural Information Processing Systems; December 2017; Long Beach, California, USA p. 4768-4777.
29. Mao Y, Wenlin C, Chen Y, Lu C, Kollef M, Bailey T. An integrated data mining approach to real-time clinical monitoring and deterioration warning. New York: Association for Computing Machinery; 2012 Presented at: KDD '12: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 2012; Beijing, China p. 1140-1148. [doi: 10.1145/2339530.2339709]
30. Sahara K, Paredes AZ, Tsilimigras DI, Sasaki K, Moro A, Hyer JM, et al. Machine learning predicts unpredicted deaths with high accuracy following hepatopancreatic surgery. Hepatobiliary Surg Nutr 2021 Jan;10(1):20-30 [FREE Full text] [doi: 10.21037/hbsn.2019.11.30] [Medline: 33575287]
31. Fest J, Ruiter TR, Groot Koerkamp B, Rizopoulos D, Ikram MA, van Eijck CHJ, et al. The neutrophil-to-lymphocyte ratio is associated with mortality in the general population: The Rotterdam Study. Eur J Epidemiol 2019 May;34(5):463-470 [FREE Full text] [doi: 10.1007/s10654-018-0472-y] [Medline: 30569368]
32. Felicetta JV. Age-related changes in calcium metabolism. Why they occur and what can be done. Postgrad Med 1989 Mar;85(4):85-6, 89. [doi: 10.1080/00325481.1989.11700616] [Medline: 2648367]
33. Ngo-Huang A, Parker NH, Wang X, Petzel MQB, Fogelman D, Schadler KL, et al. Home-based exercise during preoperative therapy for pancreatic cancer. Langenbecks Arch Surg 2017 Dec;402(8):1175-1185. [doi: 10.1007/s00423-017-1599-0] [Medline: 28710540]
34. Kleckner IR, Kamen C, Gewandter JS, Mohile NA, Heckler CE, Culakova E, et al. Effects of exercise during chemotherapy on chemotherapy-induced peripheral neuropathy: a multicenter, randomized controlled trial. Support Care Cancer 2018 Apr;26(4):1019-1028 [FREE Full text] [doi: 10.1007/s00520-017-4013-0] [Medline: 29243164]
35. Hughes MJ, Hackney RJ, Lamb PJ, Wigmore SJ, Christopher Deans DA, Skipworth RJE. Prehabilitation Before Major Abdominal Surgery: A Systematic Review and Meta-analysis. World J Surg 2019 Jul;43(7):1661-1668. [doi: 10.1007/s00268-019-04950-y] [Medline: 30788536]

Abbreviations

ACS-NSQIP SRC: American College of Surgeons National Surgical Quality Improvement Program surgical risk calculator
ANOVA: analysis of variance
AUROC: area under the receiver operating characteristic
DFA: detrended fluctuation analysis
GBT: gradient boosted trees
IRB: institutional review board
KNN: k-nearest neighbors
LR: logistic regression
LOSO CV: leave-one-subject-out cross-validation
MAGS: Modified Accordion Grading System
MESOR: midline estimating statistic of rhythm
SHAP: SHapley Additive exPlanations
SVM: support vector machine

Edited by R Kukafka; submitted 17.08.20; peer-reviewed by C Low, M Asgari Mehrabadi; comments to author 30.09.20; revised version received 18.11.20; accepted 17.02.21; published 18.03.21.

Please cite as:
Cos H, Li D, Williams G, Chininis J, Dai R, Zhang J, Srivastava R, Raper L, Sanford D, Hawkins W, Lu C, Hammill CW
Predicting Outcomes in Patients Undergoing Pancreatectomy Using Wearable Technology and Machine Learning: Prospective Cohort Study
J Med Internet Res 2021;23(3):e23595
URL: https://www.jmir.org/2021/3/e23595
doi:10.2196/23595
PMID:33734096

Original Paper

Acceptance of Telerheumatology by Rheumatologists and General Practitioners in Germany: Nationwide Cross-sectional Survey Study

Felix Muehlensiepen[1,2,3], MPH, Dr rer medic; Johannes Knitza[4], MHBA, Dr med; Wenke Marquardt[1], MSc; Jennifer Engler[5], MPH, MA, Dr rer medic; Axel Hueber[6], Dr med, PhD; Martin Welcker[7], Dr med

[1]KV Consult- und Managementgesellschaft mbH, Potsdam, Germany

[2]Center for Health Services Research, Brandenburg Medical School Theodor Fontane, Rüdersdorf bei Berlin, Germany

[3]Faculty of Health Sciences Brandenburg, Brandenburg Medical School Theodor Fontane, Neuruppin, Germany

[4]Department of Internal Medicine 3 - Rheumatology and Immunology, Friedrich-Alexander University Erlangen- Nürnberg and Universitätsklinikum Erlangen, Erlangen, Germany

[5]Institut für Allgemeinmedizin, Goethe-Universität Frankfurt am Main, Frankfurt am Main, Germany

[6]Sektion Rheumatologie, Sozialstiftung Bamberg, Bamberg, Germany

[7]Medizinisches Versorgungszentrum für Rheumatologie Dr M Welcker GmbH, Planegg, Germany

Corresponding Author:
Felix Muehlensiepen, MPH, Dr rer medic
Center for Health Services Research
Brandenburg Medical School Theodor Fontane
Seebad 82/83
Rüdersdorf bei Berlin, 15562
Germany
Phone: 49 33638 83501
Email: Felix.Muehlensiepen@mhb-fontane.de

Abstract

Background: The worldwide burden of musculoskeletal diseases is increasing. The number of newly registered rheumatologists has stagnated. Primary care, which takes up a key role in early detection of rheumatic disease, is working at full capacity. COVID-19 and its containment impede rheumatological treatment. Telemedicine in rheumatology (telerheumatology) could support rheumatologists and general practitioners.

Objective: The goal of this study was to investigate acceptance and preferences related to the use of telerheumatology care among German rheumatologists and general practitioners.

Methods: A nationwide, cross-sectional, self-completed, paper-based survey on telerheumatology care was conducted among outpatient rheumatologists and general practitioners during the pre-COVID-19 period.

Results: A total of 73.3% (349/476) of survey participants rated their knowledge of telemedicine as unsatisfactory, poor, or very poor. The majority of survey participants (358/480, 74.6%) answered that they do not currently use telemedicine, although 62.3% (291/467) would like to. Barriers to the implementation of telemedicine include the purchase of technology equipment (182/292, 62.3%), administration (181/292, 62.0%), and poor reimbursement (156/292, 53.4%). A total of 69.6% (117/168) of the surveyed physicians reckoned that telemedicine could be used in rheumatology. Surveyed physicians would prefer to use telemedicine to communicate directly with other physicians (370/455, 81.3%) than to communicate with patients (213/455, 46.8%). Among treatment phases, 64.4% (291/452) of participants would choose to use telemedicine during *follow-up*. Half of the participants would choose *telecounseling* as a specific approach to improve rheumatology care (91/170, 53.5%).

Conclusions: Before COVID-19 appeared, our results indicated generally low use but high acceptance of the implementation of telerheumatology among physicians. Participants indicated that the lack of a structural framework was a barrier to the effective implementation of telerheumatology. Training courses should be introduced to address the limited knowledge on the part of physicians in the use of telemedicine. More research into telerheumatology is required. This includes large-scale randomized controlled trials, economic analyses, and the exploration of user preferences.

(J Med Internet Res 2021;23(3):e23742) doi:10.2196/23742

KEYWORDS

telemedicine; eHealth; mHealth; rheumatology; primary care; health services research; COVID-19

Introduction

The worldwide burden of musculoskeletal diseases is increasing [1]. Growing life expectancy, widespread overweight, and a frequent lack of exercise have caused a surge in musculoskeletal disorders. Besides individual health burden, these chronic diseases create a considerable financial burden for society overall, as patients often take sick leave and early retirement [2]. While increasingly effective treatments have been developed and implemented, the number of newly registered rheumatologists has stagnated [3], and the global need for rheumatologists cannot be met [4]. General practitioners are usually the first point of contact for patients and play an important role in the early detection of rheumatic and musculoskeletal diseases (RMDs). But primary care is also affected by a shortage of staff and, in view of demographic change, an increasingly demanding work burden [5,6]. A lack of physicians has led to diagnostic delays for various rheumatologic diseases [7] and a decrease in treatment efficacy [8].

In recent decades, information and communication technologies have entered health care. One field of application is telemedicine, which is defined as follows [9]:

> *Telemedicine is the practice of medicine over a distance, in which interventions, diagnoses, therapeutic decisions, and subsequent treatment recommendations are based on patient data, documents and other information transmitted through telecommunication systems.*

The use of telemedicine in the provision of rheumatology care (telerheumatology) could ease constraints on health care access and the timeliness of care, bridge the workforce gap, and improve access to care for underserved communities [10,11]. The effectiveness of telerheumatology could be equal or higher than a standard face-to-face approach [12-16]. However, evidence is lacking, and further studies are needed to determine the best use of telerheumatology [16,17]. The current global COVID-19 pandemic has highlighted the need for telerheumatology and could promote the use of innovative solutions in clinical routines [18]. However, the effective implementation of telerheumatology requires acceptance by potential users.

The aim of this user-centered study was to assess the acceptance by physicians of the implementation of telerheumatology and to identify potential application areas in preparation for the development of telemedical approaches.

Methods

This paper reports on findings from the analysis of data collected as part of a cross-sectional, self-completed, paper-based survey of German outpatient rheumatologists and general practitioners from September to November 2018; the survey investigated acceptance, opportunities, and obstacles to the implementation of telemedicine.

The inclusion criteria for participants were that they had to be (1) rheumatologists or general practitioners, (2) practicing in ambulatory health care, (3) based in Germany, and (4) active (ie, not retired and not in training).

Two health care researchers (FM and WM) and one experienced rheumatologist (MW) designed the questionnaire. It was pilot-tested on 5 rheumatologists and 5 general practitioners to gauge the need to refine wording and format, and to check whether predefined response options were exhaustive. Minor revisions were made accordingly.

The questionnaire comprised 25 questions and was divided into three mandatory sections:

1. Telemedicine: knowledge and use.
2. Telerheumatology.
3. Sociodemographic data.

Response categories were nominal or ordinal. The questionnaire also contained several open questions.

The *telemedicine*: *knowledge and use* section included questions about expertise and the use of telemedicine. Participants were asked to rate their knowledge of telemedicine on a 6-point Likert scale. Furthermore, inquiries were made into current telemedicine use, willingness to use telemedicine, and barriers to its use.

The *telerheumatology* section included questions about the perceived usefulness of telerheumatology. This section also included questions about preferences for particular uses of telerheumatology, such as patient groups, application areas, and specific tools.

Questions about sociodemographic and occupational characteristics included age, gender, medical specialty, clinical location, type of practice, and number of patients (quarterly).

The survey was sent to all rheumatologists (n=49) and general practitioners (n=1820) in the federal state of Brandenburg, Germany, and rheumatologists (n=39) and general practitioners (n=487) in a nationwide reference group. The contact details of potential participants in Brandenburg were provided by the Brandenburg Association of Statutory Health Insurance Physicians. The nationwide reference group consisted of physicians from RHADAR (RheumaDatenRhePort GbR) and cooperating rheumatologists and general practitioners. RHADAR is an association of more than 25 rheumatologists that share the aim of developing and improving digital technology. The association also collects and analyzes anonymized patient data to identify supply and demand for rheumatology services. In total, 88 rheumatologists and 2307 general practitioners were contacted by mail.

Statistical analyses were performed using SPSS Statistics for Windows, version 22.0 (IBM Corp). Descriptive statistics included quantities, percentages, median scores, and ranges for ordinal variables.

The study was conducted in compliance with current data protection regulations and the Helsinki Declaration. All study participants were informed about the research project. Data were anonymized before analysis. The ethics committee of the Theodor Fontane Medical School in Brandenburg stated that no written consent was necessary due to the noninterventional study design.

Results

Overview

From September to November 2018, a cross-sectional, self-completed, paper-based survey on telerheumatology was filled in by German outpatient rheumatologists and general practitioners. Of the 2395 questionnaires that were sent out, 497 (20.8%) were returned. Of the 497 responses, 12 (2.4%) were excluded from the analysis because fewer than half the questions were answered. The final response rate for rheumatologists was 55% (48/88) and for general practitioners it was 18.9% (437/2307).

Sample Characteristics

Data for this study were obtained from 485 physicians (437/485, 90.1% general practitioners; 48/485, 9.9% rheumatologists) (see Table 1). About half of the respondents were between 50 and 60 years old (228/474, 48.1%). Slightly more than half of the participants were women (254/470, 54.0%). One-third of the surveyed physicians worked in a town (154/474, 32.5%), one-third worked in a provincial town (158/474, 33.3%), 17.9% (85/474) worked in a city, and 16.2% (77/474) worked in a rural area. Overall, 53.5% (252/471) of the physicians worked in a single-handed practice and 46.5% (219/471) worked in a group practice. Almost two-thirds of the surveyed physicians treated over 4000 patients per year.

Table 1. Participant demographics.

Demographic	Rheumatologists, n (%)[a]	General practitioners, n (%)[a]	*P* value	Total, n (%)[a]
Age (years)				
Total	47 (100)	427 (100)	.22	474 (100)
<40	4 (9)	36 (8.4)		40 (8.4)
40-50	16 (34)	85 (19.9)		101 (21.3)
51-60	20 (43)	208 (48.7)		228 (48.1)
>60	7 (15)	98 (23.0)		105 (22.2)
Gender				
Total	46 (100)	424 (100)	.17	470 (100)
Female	19 (41)	235 (55.4)		254 (54.0)
Male	27 (59)	189 (44.6)		216 (46.0)
Clinical location				
Total	47 (100)	427 (100)	<.001	474 (100)
City	23 (49)	62 (14.5)		85 (17.9)
Town	10 (21)	144 (33.7)		154 (32.5)
Provincial town	14 (30)	144 (33.7)		158 (33.3)
Rural area	0 (0)	77 (18.0)		77 (16.2)
Type of practice				
Total	47 (100)	424 (100)	<.001	471 (100)
Single-handed	10 (21)	242 (57.1)		252 (53.5)
Group	37 (79)	182 (42.9)		219 (46.5)
No. of patients treated (quarterly)				
Total	45 (100)	411 (100)	.14	456 (100)
<500	4 (9)	15 (3.6)		19 (4.2)
500-1000	17 (38)	130 (31.6)		147 (32.2)
>1000	24 (53)	266 (64.7)		290 (63.6)

[a]Percentages may not add up to 100% due to rounding.

Telemedicine: Knowledge and Use

A total of 73.3% (349/476) of respondents rated their knowledge of telemedicine as 4 (unsatisfactory), 5 (poor), or 6 (very poor). The minority (127/476, 26.7%) rated their knowledge of telemedicine as 1 (very good), 2 (good), or 3 (satisfactory). The majority (358/480, 74.6%) did not currently use telemedicine, but 62.3% (291/467) answered that they would like to use it (see Table 2). A total of 89.3% (259/290) of surveyed physicians indicated that barriers prevented them from using telemedicine. The top three obstacles to the introduction of telemedicine according to respondents were the purchase of technology equipment (182/292, 62.3%), administration (181/292, 62.0%), and inadequate remuneration (156/292, 53.4%).

Table 2. Telemedicine: knowledge and use.

Question and responses	Rheumatologists, n (%)[a]	General practitioners, n (%)[a]	*P* value	Total, n (%)[a]
How do you rate your own knowledge of telemedicine?				
Total	47 (100)	429 (100)	.14	476 (100)
1 (very good)	1 (2)	19 (4.4)		20 (4.2)
2 (good)	8 (17)	34 (7.9)		42 (8.8)
3 (satisfactory)	9 (19)	56 (13.1)		65 (13.7)
4 (unsatisfactory)	9 (19)	110 (25.6)		119 (25.0)
5 (poor)	16 (34)	138 (32.2)		154 (32.4)
6 (very poor)	4 (9)	72 (16.8)		76 (16.0)
Do you use telemedicine?				
Total	47 (100)	433 (100)	.15	480 (100)
Yes	16 (34)	106 (24.5)		122 (25.4)
No	31 (66)	327 (75.5)		358 (74.6)
Would you like to use telemedicine?				
Total	46 (100)	421 (100)	.45	467 (100)
Yes	31 (67)	260 (61.8)		291 (62.3)
No	15 (33)	161 (38.2)		176 (37.7)
Does anything prevent you from using telemedicine?				
Total	30 (100)	260 (100)	.26	290 (100)
Yes	25 (83)	234 (90.0)		259 (89.3)
No	5 (17)	26 (10.0)		31 (10.7)
What prevents you from using telemedicine? (multiple selections possible)				
Total	31 (100)	261 (100)		292 (100)
Purchase of technology equipment	16 (52)	166 (63.6)	.19	182 (62.3)
Administration	21 (68)	160 (61.3)	.49	181 (62.0)
Poor reimbursement	21 (68)	135 (51.7)	.09	156 (53.4)
Data security	15 (48)	120 (46.0)	.80	135 (46.2)
Lack of participation by colleagues	8 (26)	89 (34.1)	.35	97 (33.2)
Technical comprehension of patients	12 (39)	84 (32.3)	.47	96 (32.9)
Poor internet connection	5 (16)	52 (19.9)	.61	57 (19.5)

[a]Percentages may not add up to 100% due to rounding or where multiple selections were possible.

Implementation of Telerheumatology

A total of 69.6% (117/168) of participants considered telemedicine as usable in rheumatology (see Table 3).

When asked who should interact using telemedicine, 81.3% (370/455) responded *physician-physician*, 46.8% (213/455) responded *physician-patient*, and 25.7% (117/455) responded *physician-assistant* (multiple replies were possible). The preferred therapy phases for the use of telemedicine were *follow-up* (291/452, 64.4%), *initial contact* (153/452, 33.8%), and *screening* (133/452, 29.4%). Participants were asked to indicate specific digital tools that could support rheumatology care. The most frequently selected items were *telecounseling* (91/170, 53.5%), *telediagnostics* (76/170, 44.7%), and *video consultations* (67/170, 39.4%). These were followed by *online*

appointment assignments (56/170, 32.9%), *e-learning* (55/170, 32.4%), *patient apps* (48/170, 28.2%), *digital screening* (36/170, 21.2%), *wearable devices* (20/170, 11.8%), *telesurgery* (7/170, 4.1%), and *other tools* (4/170, 2.4%).

Table 3. Implementation of telemedicine in rheumatology care.

Question and responses	Rheumatologists, n (%)[a]	General practitioners, n (%)[a]	*P* value	Total, n (%)[a]
Is telemedicine usable in rheumatology?				
Total	43 (100)	125 (100)	.18	168 (100)
Yes	32 (74)	85 (68.0)		117 (69.6)
No	11 (26)	40 (32.0)		51 (30.4)
Which parties should establish communication via telemedicine? (multiple selections possible)				
Total	45 (100)	410 (100)		455 (100)
Physician-physician	36 (80)	334 (81.5)	.88	370 (81.3)
Physician-patient	28 (62)	185 (45.1)	.03	213 (46.8)
Physician-assistant	14 (31)	103 (25.1)	.38	117 (25.7)
Other participants and combinations	5 (11)	13 (3.2)	.009	18 (4.0)
No communication	3 (7)	54 (13.2)	.21	57 (12.6)
At which stages can telemedicine support rheumatology care? (multiple selections possible)				
Total	45 (100)	407 (100)		452 (100)
Screening	23 (51)	110 (27.0)	.001	133 (29.4)
Initial contact	11 (24)	142 (34.9)	.16	153 (33.8)
Follow-up	32 (71)	259 (63.6)	.32	291 (64.4)
Other stages	7 (16)	38 (9.4)	.19	45 (10.0)
At no stage	8 (18)	69 (17.0)	.90	77 (17.1)
Which tools could support rheumatologic care? (multiple selections possible)				
Total	44 (100)	126 (100)		170 (100)
Telecounseling	24 (55)	55 (43.7)	.21	79 (45.5)
Telediagnostics	18 (41)	58 (46.0)	.56	76 (44.7)
Video consultations	19 (43)	48 (38.1)	.55	67 (39.4)
Online appointment assignments	20 (45)	36 (28.6)	.04	56 (32.9)
e-Learning	15 (34)	40 (31.7)	.78	55 (32.4)
Patient apps	17 (39)	31 (24.6)	.08	48 (28.2)
Digital screening	15 (34)	21 (16.7)	.02	36 (21.2)
Wearable devices	9 (20)	11 (8.7)	.04	20 (11.8)
Telesurgery	3 (7)	4 (3.2)	.30	7 (4.1)
Other tools	2 (5)	2 (1.6)	.27	4 (2.4)
No tools	6 (14)	15 (11.9)	.76	21 (12.4)

[a]Percentages do not add up to 100% where multiple selections were possible.

Discussion

Principal Findings

To the best of our knowledge, we have performed the largest nationwide survey in Germany on the use of telemedicine in adult rheumatology aimed at promoting and guiding the implementation of telerheumatology.

Rheumatologists and general practitioners generally consider the overall application of telerheumatology to be acceptable, and two-thirds of the respondents would like to implement telemedicine in their everyday practice. Rheumatologists expressed an even greater willingness to use telemedicine than general practitioners. Rheumatologists and general practitioners welcomed a wide variety of approaches to telemedicine. However, only a minority of physicians already used telemedicine at the time the survey was conducted. Barriers to its introduction, such as limited knowledge, administrative

expenses, the purchase of technology equipment, and inadequate reimbursement, were clearly identified by specialists and generalists. The results provide information on how telemedicine can support rheumatology care from the physician's perspective. Conservative and familiar communication formats, such as the exchange of information with colleagues, were preferred to the direct exchange of information with patients and assistants. This was demonstrated by the high approval of the item *telecounseling*. Although various telecounseling tools already exist, their development for rheumatology applications is not as advanced as in, for example, intensive care and cardiology. This is reflected in the relatively small number of respondents that were using telemedicine at the time of the survey.

Limitations

The average age of our sample is similar to that of German physicians as a whole [19]. Women were slightly overrepresented compared to the average [19], which may also indicate that female physicians are more interested in telehealth. Due to the small number of rheumatologists, this survey mainly reports on the opinion of general practitioners. Although the survey was directed at physicians from all over Germany, it was primarily physicians from Brandenburg who participated due to the recruitment strategy. We assume a self-selection bias and a nonresponse bias, as the survey was probably mainly answered by physicians who are interested in telemedicine and rheumatology. To overcome this bias, we chose a paper-based survey. An online survey may have increased the response rate and reduced the effort for data management. However, we assumed an online survey would have influenced response behavior, in the sense of a positive bias toward users of telemedicine. To answer the questionnaire, knowledge in the field of telemedicine was required, as, for example, preferences for specific tools were asked about. Considering the limited knowledge of doctors in the field of telemedicine, response biases are, therefore, likely. Furthermore, we assume that due to rapid technical developments in the field, our predefined response categories were not exhaustive. Only ambulatory physicians were asked to participate in the survey, so the results do not provide any information on the acceptance by patients, hospital staff, or other professional groups other than outpatient physicians. Lastly, the survey was conducted in pre-COVID-19 times, so further research on the development of acceptance of telerheumatology use is highly needed.

Comparison With Prior Work

This work adds to the growing body of telerheumatology knowledge [11-17] by providing detailed user preferences, needs, and barriers. Hence, we believe that the results of this study may help in the development of telemedicine solutions that can be integrated into the clinical routine of patients with rheumatic diseases.

In contrast to the results of a recent study that identified a negative attitude toward digitization in the health care system among physicians in Germany [20], our results indicated that physicians have a positive attitude toward telemedicine. An American Medical Association survey of 3500 physicians in the United States found that fewer than 10% of rheumatologists used telemedicine, which is significantly less than physicians from other medical fields, such as radiologists (43%) [21], and less than the percentage of rheumatologists that use telemedicine according to our study (34%). Although most respondents reckoned that telecounseling may support rheumatology care, telecounseling is not used or is rarely used. In a nationwide survey of digitization among German medical practices, only 1% of surveyed physicians responded that they use videoconferences to communicate with other physicians. The most frequent answers were *email* and *no digital communication at all* [22]. In the survey on digitization in German medical practices, security gaps in information technology (IT), the considerable cost and effort involved in introducing digital technologies, and an unfavorable cost-benefit ratio were identified as the main obstacles to digitization from the perspective of physicians [22]. Participants in our survey attached relatively little importance to security gaps in IT, with only 46.2% choosing *data security* as an issue that prevented them from using telemedicine. Televisual consultations with patients appear to have considerable potential in rheumatology care, and especially in initial consultations [12]. However, only a minority of participants in our survey favored the use of telemedicine in initial consultations. This finding confirms the results of a comparable study on veteran rheumatology care from the United States [23]. Furthermore, only one-third of those surveyed wished to use telemedicine in direct patient contacts at all, which contrasts starkly with overall telemedical developments in the health care system [16,17]. This is unfortunate, as previous research has shown that patients with rheumatoid arthritis consider telerheumatology to be a flexible solution that increases the independence of health authorities and raises personal knowledge [24]. Other studies indicate that health care resulting from televisits is as effective as that provided following in-person visits [25,26]. A qualitative study also reports that patients would be willing to accept electronic collection and sharing of patient-reported outcomes (PROs) between clinical encounters if it facilitated communication with health care providers and provided access to reliable information on RMDs [27]. However, a recent study has shown that rheumatologists are reluctant to study electronic PROs because it would lead to a massive increase in their workload [28].

Mobile apps promise to accelerate diagnostic investigations and improve monitoring and the outcomes of patients with RMDs [29]. The small number of rheumatologists that supported the use of apps to improve clinical routines (34%) contrasts with previous research from 2018 in which 49% said they were already using such apps [30]. One of the top reasons for the reluctance of rheumatologists to use apps may be lack of evidence [31]. The combination of electronic PROs [28] and objective serology and genetic testing promises to result in the earlier diagnosis of RMDs [32].

Our findings indicate that rheumatologists accept telemedicine to a greater extent than general practitioners. Furthermore, the preferences of rheumatologists differ from those of general practitioners with regard to which specific tools could be implemented into rheumatology care and when. These variations may be related to differences in the two professions as well as the distinct phases of the disease in which rheumatologists and general practitioners encounter RMDs. We also analyzed

differences in the acceptance and preferences of telemedicine with regard to the age and gender of the physicians as well as the type and region of their practices. No differences or only small differences were found.

Perspectives for Telerheumatology

COVID-19 has demonstrated the importance of contactless approaches in medical care. As early as 2018, when we conducted the survey, rheumatologists and general practitioners were willing to use telerheumatology. It is assumed that this readiness has increased due to the pandemic, which is likely to strongly accelerate the use of telemedicine as society adopts new health care standards [18].

Nevertheless, the great potential of telemedicine is not being fully reached. Further research into the implementation of telerheumatology is desperately needed. This includes large-scale randomized controlled studies on the effects and health economic outcomes, as well as risks and adverse events, of specific interventions.

As our results indicate that there will be no "one-size-fits-all" solution in the realm of telemedicine, further research into the perspectives and preferences of physicians, patients, and other telemedicine users in rheumatology is essential. This may provide the foundation for individual patient- and physician-adapted telemedicine options and triage mechanisms to select patients for either digital or analog consultations, as appropriate [25,26].

As physicians reported barriers to the use of telemedicine, it would appear that the structural framework is not yet in place for the effective implementation of telerheumatology. Considerable administrative effort and inadequate reimbursement structures prevented the surveyed physicians from using telemedicine. However, the greatest barrier was the limited knowledge on the part of physicians on the use of telemedicine, which highlights the need for the timely introduction of low-threshold training courses.

Conclusions

Our study showed that rheumatologists and general practitioners support the implementation of telerheumatology, and two-thirds of respondents would like to implement telemedicine into their clinical routine. Rheumatologists expressed an even greater willingness to use telemedicine than general practitioners, with respondents welcoming a variety of telemedicine approaches. However, only a minority of the surveyed physicians currently use telemedicine. Furthermore, most physicians regard their knowledge of telemedicine as rather poor. The provision of high-quality rheumatology care using telemedicine will require urgently needed research as well as a reduction in existing barriers and training for specialists and generalists.

Acknowledgments

The authors would like to thank the participants, their teams, and all other supporters of this study. We also owe special gratitude to KV Consult- und Managementgesellschaft mbH, which initiated the study. We thank Phillip Elliot for editing the manuscript. Financial support was provided by Novartis Pharma GmbH. Novartis had no role in the study design or in the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication. Publication of this article was not contingent upon approval by Novartis.

Authors' Contributions

All authors were involved in drafting the article and critically revising it for important intellectual content, and all authors approved the final version to be submitted for publication. FM had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. FM, WM, and MW were responsible for the study conception and design and for acquisition of the data. FM, JK, WM, JE, AH, and MW were responsible for analysis and interpretation of the data.

Conflicts of Interest

None declared.

References

1. Sebbag E, Felten R, Sagez F, Sibilia J, Devilliers H, Arnaud L. The world-wide burden of musculoskeletal diseases: A systematic analysis of the World Health Organization Burden of Diseases Database. Ann Rheum Dis 2019 Jun;78(6):844-848. [doi: 10.1136/annrheumdis-2019-215142] [Medline: 30987966]
2. Sullivan PW, Ghushchyan V, Huang X, Globe DR. Influence of rheumatoid arthritis on employment, function, and productivity in a nationally representative sample in the United States. J Rheumatol 2010 Mar;37(3):544-549. [doi: 10.3899/jrheum.081306] [Medline: 20080920]
3. Krusche M, Sewerin P, Kleyer A, Mucke J, Vossen D, Morf H, Rheumadocs und Arbeitskreis Junge Rheumatologie (AGJR). Specialist training quo vadis? [Article in German]. Z Rheumatol 2019 Oct;78(8):692-697. [doi: 10.1007/s00393-019-00690-5] [Medline: 31468164]
4. Al Maini M, Adelowo F, Al Saleh J, Al Weshahi Y, Burmester G, Cutolo M, et al. The global challenges and opportunities in the practice of rheumatology: White paper by the World Forum on Rheumatic and Musculoskeletal Diseases. Clin Rheumatol 2015 May;34(5):819-829 [FREE Full text] [doi: 10.1007/s10067-014-2841-6] [Medline: 25501633]

5. Hobbs FDR, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, National Institute for Health Research School for Primary Care Research. Clinical workload in UK primary care: A retrospective analysis of 100 million consultations in England, 2007-14. Lancet 2016 Jun 04;387(10035):2323-2330 [FREE Full text] [doi: 10.1016/S0140-6736(16)00620-6] [Medline: 27059888]
6. Ulrich L, Lemke D, Erler A, Müller B, Gerlach F, Dahlhaus A. Work stress among family physicians in two Thuringian districts [Article in German]. Z Allg Med 2019;95:112-118 [FREE Full text] [doi: 10.3238/zfa.2019.0112-0118]
7. Sørensen J, Hetland ML. Diagnostic delay in patients with rheumatoid arthritis, psoriatic arthritis and ankylosing spondylitis: Results from the Danish nationwide DANBIO registry. Ann Rheum Dis 2015 Mar;74(3):e12 [FREE Full text] [doi: 10.1136/annrheumdis-2013-204867] [Medline: 24534758]
8. Quinn M, Emery P. Window of opportunity in early rheumatoid arthritis: Possibility of altering the disease process with early intervention. Clin Exp Rheumatol 2003;21(5 Suppl 31):S154-S157 [FREE Full text] [Medline: 14969068]
9. WMA Statement on the Ethics of Telemedicine. World Medical Association. 2018. URL: https://www.wma.net/policies-post/wma-statement-on-the-ethics-of-telemedicine/ [accessed 2021-03-05]
10. Ward IM, Schmidt TW, Lappan C, Battafarano DF. How critical is tele-medicine to the rheumatology workforce? Arthritis Care Res (Hoboken) 2016 Oct;68(10):1387-1389 [FREE Full text] [doi: 10.1002/acr.22853] [Medline: 26866514]
11. Rezaian MM, Brent LH, Roshani S, Ziaee M, Sobhani F, Dorbeigi A, et al. Rheumatology care using telemedicine. Telemed J E Health 2020 Mar;26(3):335-340. [doi: 10.1089/tmj.2018.0256] [Medline: 31084537]
12. Leggett P, Graham L, Steele K, Gilliland A, Stevenson M, O'Reilly D, et al. Telerheumatology--Diagnostic accuracy and acceptability to patient, specialist, and general practitioner. Br J Gen Pract 2001 Sep;51(470):746-748 [FREE Full text] [Medline: 11593837]
13. Kuusalo L, Sokka-Isler T, Kautiainen H, Ekman P, Kauppi MJ, Pirilä L, SandRA Study Group. Automated text message-enhanced monitoring versus routine monitoring in early rheumatoid arthritis: A randomized trial. Arthritis Care Res (Hoboken) 2020 Mar;72(3):319-325. [doi: 10.1002/acr.23846] [Medline: 30740935]
14. Salaffi F, Carotti M, Ciapetti A, Di Carlo M, Gasparini S, Farah S, et al. Effectiveness of a telemonitoring intensive strategy in early rheumatoid arthritis: Comparison with the conventional management approach. BMC Musculoskelet Disord 2016 Apr 02;17:146 [FREE Full text] [doi: 10.1186/s12891-016-1002-2] [Medline: 27038788]
15. de Thurah A, Stengaard-Pedersen K, Axelsen M, Fredberg U, Schougaard LMV, Hjollund NHI, et al. Tele-health followup strategy for tight control of disease activity in rheumatoid arthritis: Results of a randomized controlled trial. Arthritis Care Res (Hoboken) 2018 Mar;70(3):353-360. [doi: 10.1002/acr.23280] [Medline: 28511288]
16. Piga M, Cangemi I, Mathieu A, Cauli A. Telemedicine for patients with rheumatic diseases: Systematic review and proposal for research agenda. Semin Arthritis Rheum 2017 Aug;47(1):121-128. [doi: 10.1016/j.semarthrit.2017.03.014] [Medline: 28420491]
17. McDougall JA, Ferucci ED, Glover J, Fraenkel L. Telerheumatology: A systematic review. Arthritis Care Res (Hoboken) 2017 Oct;69(10):1546-1557 [FREE Full text] [doi: 10.1002/acr.23153] [Medline: 27863164]
18. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
19. Statistical Information From the Federal Register of Physicians [Report in German]. Berlin, Germany: Kassenärztliche Bundesvereinigung; 2020. URL: https://www.kbv.de/media/sp/2020-12-31_BAR_Statistik.pdf [accessed 2021-03-23]
20. Hansen A, Herrmann M, Ehlers JP, Mondritzki T, Hensel KO, Truebel H, et al. Perception of the progressing digitization and transformation of the German health care system among experts and the public: Mixed methods study. JMIR Public Health Surveill 2019 Oct 28;5(4):e14689 [FREE Full text] [doi: 10.2196/14689] [Medline: 31661082]
21. Kane CK, Gillis K. The use of telemedicine by physicians: Still the exception rather than the rule. Health Aff (Millwood) 2018 Dec;37(12):1923-1930. [doi: 10.1377/hlthaff.2018.05077] [Medline: 30633670]
22. Praxisbarometer Digitalisierung 2019. Stand und Perspektiven der Digitalisierung in der vertragsärztlichen und -psychotherapeutischen Versorgung [Report in German]. Berlin, Germany: Kassenärztliche Bundesvereinigung; 2019. URL: https://www.kbv.de/media/sp/KBV_Praxisbarometer_Digitalisierung_2019.pdf [accessed 2021-03-19]
23. Matsumoto RA, England BR, Mastarone G, Richards JS, Chang E, Wood PR, et al. Rheumatology clinicians' perceptions of telerheumatology within the veterans health administration: A national survey study. Mil Med 2020 Dec 30;185(11-12):e2082-e2087. [doi: 10.1093/milmed/usaa203] [Medline: 32789463]
24. Knudsen LR, de Thurah A, Lomborg K. Experiences with telehealth followup in patients with rheumatoid arthritis: A qualitative interview study. Arthritis Care Res (Hoboken) 2018 Sep;70(9):1366-1372. [doi: 10.1002/acr.23483] [Medline: 29195011]
25. Kulcsar Z, Albert D, Ercolano E, Mecchella JN. Telerheumatology: A technology appropriate for virtually all. Semin Arthritis Rheum 2016 Dec;46(3):380-385. [doi: 10.1016/j.semarthrit.2016.05.013] [Medline: 27395561]
26. Taylor-Gjevre R, Nair B, Bath B, Okpalauwaekwe U, Sharma M, Penz E, et al. Addressing rural and remote access disparities for patients with inflammatory arthritis through video-conferencing and innovative inter-professional care models. Musculoskeletal Care 2017 Oct 02;16(1):90-95. [doi: 10.1002/msc.1215] [Medline: 28967235]

27. Navarro-Millán I, Zinski A, Shurbaji S, Johnson B, Fraenkel L, Willig J, et al. Perspectives of rheumatoid arthritis patients on electronic communication and patient-reported outcome data collection: A qualitative study. Arthritis Care Res (Hoboken) 2019 Jan;71(1):80-87 [FREE Full text] [doi: 10.1002/acr.23580] [Medline: 29669191]
28. Krusche M, Klemm P, Grahammer M, Mucke J, Vossen D, Kleyer A, et al. Acceptance, usage, and barriers of electronic patient-reported outcomes among German rheumatologists: Survey study. JMIR Mhealth Uhealth 2020 Jul 20;8(7):e18117 [FREE Full text] [doi: 10.2196/18117] [Medline: 32390592]
29. Krusche M, Ruffer N, Grahammer M, Knitza J. Mobile apps and their usage in rheumatology [Article in German]. Z Rheumatol 2020 Aug;79(6):554-561. [doi: 10.1007/s00393-020-00822-2] [Medline: 32472178]
30. Knitza J, Vossen D, Geffken I, Krusche M, Meyer M, Sewerin P, Arbeitskreis Junge Rheumatologen. Use of medical apps and online platforms among German rheumatologists: Results of the 2016 and 2018 DGRh conference surveys and research conducted by rheumadocs [Article in German]. Z Rheumatol 2019 Nov;78(9):839-846. [doi: 10.1007/s00393-018-0578-3] [Medline: 30542914]
31. Knitza J, Tascilar K, Messner E, Meyer M, Vossen D, Pulla A, et al. German mobile apps in rheumatology: Review and analysis using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2019 Aug 05;7(8):e14991 [FREE Full text] [doi: 10.2196/14991] [Medline: 31381501]
32. Knitza J, Knevel R, Raza K, Bruce T, Eimer E, Gehring I, JPAST Group. Toward earlier diagnosis using combined eHealth tools in rheumatology: The Joint Pain Assessment Scoring Tool (JPAST) project. JMIR Mhealth Uhealth 2020 May 15;8(5):e17507 [FREE Full text] [doi: 10.2196/17507] [Medline: 32348258]

Abbreviations

IT: information technology
PRO: patient-reported outcome
RHADAR: RheumaDatenRhePort GbR
RMD: rheumatic and musculoskeletal disease
Telerheumatology: telemedicine in rheumatology

Edited by C Basch, R Kukafka; submitted 21.08.20; peer-reviewed by P Sewerin, AS Aslam; comments to author 13.10.20; accepted 04.03.21; published 29.03.21.

Please cite as:
Muehlensiepen F, Knitza J, Marquardt W, Engler J, Hueber A, Welcker M
Acceptance of Telerheumatology by Rheumatologists and General Practitioners in Germany: Nationwide Cross-sectional Survey Study
J Med Internet Res 2021;23(3):e23742
URL: https://www.jmir.org/2021/3/e23742
doi:10.2196/23742
PMID:33690147

Original Paper

Association of Mobile Instant Messaging Chat Group Participation With Family Functioning and Well-Being: Population-Based Cross-sectional Study

Sheng Zhi Zhao[1], MPH; Tzu Tsun Luk[1], PhD; Ningyuan Guo[1], BSc; Man Ping Wang[1], PhD; Agnes Yuen Kwan Lai[1], PhD; Bonny Yee Man Wong[2], PhD; Daniel Yee Tak Fong[1], PhD; Sophia Siu Chee Chan[1], PhD; Tai Hing Lam[2], MD

[1]School of Nursing, The University of Hong Kong, Hong Kong, China (Hong Kong)

[2]School of Public Health, The University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
Man Ping Wang, PhD
School of Nursing
The University of Hong Kong
21 Sassoon Road
Pokfulam
Hong Kong
China (Hong Kong)
Phone: 852 3917 6636
Email: mpwang@hku.hk

Abstract

Background: Convenient and quality family communication improves family functioning and well-being. Using mobile instant messaging (IM) for family communication is increasingly popular, but its association with family functioning and family well-being has not been reported.

Objective: The aim of this study was to examine the association of the use of family IM chat groups with family functioning and well-being, and the mediating effect of family communication quality among Chinese adults in Hong Kong.

Methods: We analyzed data from the Family and Health Information Trend Survey (FHInTS), a territory-wide, probability-based telephone survey conducted in 2017. The quality of family communication, family functioning, and well-being was assessed using the Family Communication Scale; Family Adaptation, Partnership, Growth, Affection, and Resolve (APGAR) Scale; and Family Well-Being Scale (family heath, harmony, and happiness), respectively. Respondents also reported the number of family IM chat groups (0, 1, 2, ≥3), and numbers of IM messages received (<1, 1-2, 3-10, 11-20, >20) and sent (<1, 1-2, 3-10, 11-20, >20) daily. The frequency of family IM chat interaction (range 0-8) was calculated by combining the number of messages received from and sent to the family IM chat groups daily. Covariates included sociodemographic characteristics and the frequency of family face-to-face communication (often, sometimes, seldom, or never). Data were weighted by sex, age, and education of the general population. Adjusted β coefficients of family functioning and well-being in relation to having a family IM chat group, and numbers of messages received and sent were calculated. The mediation effects of family communication on these associations were assessed, controlling for the covariates.

Results: A random sample of 1638 Chinese adults (45.6% men; 78.1% aged 25 to 64 years) were interviewed (response rate: 74.4%). Female, younger age, being married or cohabiting, higher education, higher income, better family functioning, and well-being were associated with having at least one family IM chat group (all P<.01). Higher scores of family communication, family APGAR, and family well-being were associated with having more family IM chat groups and more messages received from and sent to family IM chat groups daily (all P for trend <.01). More frequent family IM chat interaction was associated with higher scores of family communication, family APGAR, and family well-being (β=.16-.83, all P for trend <.001). The associations of family IM chat interaction with family functioning and well-being were moderately (51.0%-59.6%) mediated by family communication.

Conclusions: Use of a family IM chat group was associated with higher family functioning and well-being, and the association was partially mediated by family communication.

(J Med Internet Res 2021;23(3):e18876) doi:10.2196/18876

KEYWORDS

mobile instant messaging; chat groups; family communication; family well-being; family functioning

Introduction

Family functioning includes the interactions and relationships, adaptability, organization, and communication of the family environment, and is the foundation for the physical and psychosocial well-being of family members [1,2]. Family well-being, usually conceptualized as "family life satisfaction," is the keystone of a harmonious society [3]. From a macro perspective, social resources, including income, time, and psychological and social capital, are determinants of family functioning and well-being [4]. Individually, spending time and having quality communication with family members improves family well-being, as this allows family members to connect, achieve fulfillment, and express and share attitudes and values. Higher quality family communication improves cardiovascular health, immune system, subjective well-being, and quality of life [5,6]. Having poor family communication and relationships increase the risks of depression, anxiety, loneliness, substance abuse, and other psychological distress [7,8].

The development of information and communication technologies (ICTs) has largely transformed interpersonal communication [9]. Advanced ICT tools in smartphones, such as voice messages, multimedia messages, and video calls, have facilitated the rapid exchange of information or affect [10]. However, inappropriate use of ICTs has led to a wide array of psychological problems, including decline in the size of social circles and life satisfaction, and an increase in depression and loneliness [11]. Specifically, entertainment usage, especially gaming, is associated with emotional exhaustion, function impairment, and relationship disruption [12]. A poor perceived well-being can also result from a perceived waste of time in excessive smartphone use [13]. Relationship fulfillment is usually acquired from achieving common goals or relating to others. As the most important social tie with effort to maintain, family cohesion and satisfaction require interactions with both high quantity and quality. Under the current busy social context, communication through a smartphone, including phone calls or text messages, has been associated with increased family functioning and well-being, presumably via a positive and effective experience of interactive connection [14-16].

Face-to-face communication has remained the dominant mode of family interaction [17]. Telephone, as a traditional means of remote communication, provides instant feedback across geographical distance. Instant messaging (IM) apps (eg, WhatsApp, WeChat), with access regardless of time, distance, and cost restrictions, may function as better substitutes when face-to-face communication is hard to achieve [18]. We previously reported that more than half of the Hong Kong population used IM for family communication in the 2012 household survey [19]. A higher well-being level was observed among families using smartphones for communication [19]. Long working hours have left little time for family gatherings. The increase in smartphone-mediated communication such as mobile voice and video calls is related to closer and higher-quality relationships, which are in turn related to greater life satisfaction and well-being [20]. IM features help to maintain real-time bonding and bridge social capital, which can facilitate communication for higher perceived well-being [19,20]. A group chat allows for all sources of information to be synchronously exchanged among three or more participants, and is now commonly used among classmates, coworkers, friends, and family members [21]. As a common platform for family life information sharing, an IM chat group can strengthen family management (eg, parenting, childcare) and health information transfer for improving family health [22,23].

Positive communication experiences are critical for the adaptive functioning of family relationships by buffering against family problems and difficulties [14]. Chinese culture values family communication as an important embodiment of family cohesion [24], which is considered to be vital for the family harmony, happiness, and health (3Hs) that underlie family well-being [3,25]. Lack of positive interaction, in conjunction with a stressful work life, can be detrimental to family well-being [26]. In Hong Kong, smartphone penetration reached 88.6% in 2017 and almost all internet users (98.1%) use smartphones for communication [27]. As a new approach for family communication, there is a lack of research to determine whether a family group chat is associated with well-being. Most of the recent studies in this field have focused on IM use among teenagers and adolescents with a small sample size [28,29]; however, IM users are increasing rapidly in all age groups [27]. Therefore, the aim of this study was to examine the associations of family chat group usage with family functioning and well-being in Hong Kong adults. We also tested whether the above associations are mediated by improved quality of family communication.

Methods

Sampling

We used data from the latest phase of the Hong Kong Family and Health Information Trends Survey (FHInTS) conducted from February to May 2017 under the Hong Kong Jockey Club FAMILY project. The FHInTS was a probability-based telephone survey that aimed to assess the use of ICTs in relation to family relationships in the Hong Kong Chinese population. Details of the research design have been reported elsewhere [19,30,31]. The Public Opinion Programme (POP) of the University of Hong Kong, one of the most reputable survey agencies in Hong Kong, conducted the telephone interviews. Respondents were Hong Kong residents aged 18 years or above who could speak Cantonese. All interviews were conducted by trained interviewers using a computer-assisted telephone interviewing system. POP adopted a two-stage random sampling method. First, the residential telephone directories that covered about 76% of Hong Kong residents [32] were used to generate a list of randomized household telephone numbers (seed numbers) by a computer program. To capture the unlisted telephone numbers, plus or minus one or two of the last digits of the seed numbers were used to generate new numbers for

random dialing. Invalid household numbers and nonresponses (after 5 calls at different times and days of the week) were excluded. In the second stage, the "next birthday" rule was used to pick one eligible household resident with the closest next birthday to the interview dates as the respondent for the survey. Of the 4054 respondents who completed the interview (response rate: 74.4%), a subset of 1638 (40.4%) respondents were randomly selected to answer questions on the use of a family IM chat group, family communication, family functioning, and family well-being. Ethical approval was granted by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster. Verbal informed consent was obtained and recorded verbatim in the telephone survey.

Measurement

The interviewers read out the definition of family (ie, family members who are related through biological, marital, cohabitation, or emotional bonding) before the interview. To assess the use of a family IM chat group, the respondents were asked "Do you have a family group of 3 or more persons using IM such as WhatsApp and WeChat group chat?" with response options of "Yes" and "No." Those who responded "Yes" were further asked about the number of family IM chat groups they have (1, 2, 3, or >3) and the number of messages they receive from and send daily in the family IM chat groups separately with responses categorized into "None," "1-2 messages," "3-10 messages," "11-20 messages," and "Over 20 messages."

The quality of family communication was measured by the Family Communication Scale (FCS) [33], which consists of 10 items on a one-dimensional scale focusing on positive family communication skills between family members, such as clear and congruent messages, empathy, supportive phrases, and effective problem-solving skills. The FCS aims to gather important aspects of family communication, including the level of openness and honesty to exchange ideas, emotional tone of the interactions, and affections and concerns for each other. Each item is scored on a Likert scale of 1 (strongly disagree) to 5 (strongly agree). All items are summed to give a total score of 10 to 50, with a higher score indicating better family communication. The internal consistency of the FCS in this study was excellent (Cronbach α=.90).

The Family Adaptation, Partnership, Growth, Affection, and Resolve (APGAR) scale consists of five questions assessing the frequency of feeling satisfied with five family functioning parameters on a 3-point Likert scale of 0 (hardly ever) to 2 (almost always) [34]. The statements used in the scale concerned the emotion expression, communicative, and interactive relationships between the respondents and their family, such as "You were satisfied with the care of your family when you had different emotions (including happiness, anger, sadness)." The Chinese version of the Family APGAR has been widely used in the Hong Kong population [35] and the Cronbach α in this study was good (.86).

The Family Well-Being Scale developed based on our qualitative interviews under the FAMILY project [3,36] consists of three items of family 3Hs, each measured on a scale of 0 to 10 with a higher score indicating better family 3Hs [3,37]. A composite score of family well-being (range 0-10) was the average of the family 3Hs score. The internal consistency of the 4 items was good with a Cronbach α of .89 in this study.

Sex, age, marital status (never been married, married or cohabitating, divorced or separated, and widowed) and socioeconomic status (SES), including education attainment (primary or below, secondary, or tertiary or above) and monthly household income (in Hong Kong dollars [HK $], in which US $1=HK $7.80: ≤9,999, 10,000-19,999, 20,000-29,999, 30,000-39,999, ≥40,000, and unstable) of the respondents were recorded. We also collected data on the frequency of face-to-face family communication (often, sometimes, seldom, or never).

Statistical Analysis

The χ^2 test was used to compare sociodemographic differences and family relationship measurements between family chat group users and nonusers. To improve the representativeness of the sample, all data were weighted according to provisional figures obtained from the Census and Statistics Department on the sex-age distributions of the Hong Kong general population in 2015 and the education attainment distribution in 2011. The mean scores of the FCS, Family APGAR, and Family Well-Being Scale were compared between respondents with or without family IM chat groups and according to usage of the family IM chat group. Multivariable linear regressions were used to calculate the regression coefficient (β) of family communication, family functioning, and family well-being in relation to the use of the family IM chat group, adjusting for age, sex, marital status, and SES (model 1). In model 2, we additionally adjusted for the frequency of face-to-face family communication. We created a composite variable for the frequency of family IM chat interaction (range 0-8) by combining the variable of the number of messages received from and sent to the family IM chat groups daily for a mediation test. The Baron and Kenny approach was used to examine the mediating effect of family communication on the associations of family IM chat interaction with family functioning and well-being [38]. The Sobel test was used to determine whether the mediating (indirect) effect was significant. Bootstrapping with 1000 replications was used to calculate the 95% CI of the indirect effect. All analyses were performed using STATA version 15.0 (StataCorp LP, College Station, TX, USA). A two-sided *P* value<.05 was considered statistically significant.

Results

Of the 1638 respondents, just under half were men and the majority were aged 25 to 64 years. Most of the respondents were married or cohabitating, attained secondary or higher level of education, and had a monthly household income of more than HK $30,000 (Table 1). Almost all respondents communicate with family members (98.6%, 1341/1360), with the great majority often communicating face-to-face. Bivariate analysis showed that respondents of female sex, younger age, married or cohabitating, with higher education, higher income, better family functioning, and better well-being were associated with having at least one family IM chat group (all *P*<.01). The frequency of family face-to-face communication was not associated with the use of family IM chat groups (*P*=.14).

Most respondents (72.0%, 1180/1638) had at least one family IM chat group. Among those who have family IM chat groups, 92.0% (1083/1177) and 83.7% (987/1179) of the respondents received and sent at least one message daily, respectively (Table 2). Compared with having no family IM chat groups (n=458), having 3 family IM chat groups was associated with higher mean scores (Multimedia Appendix 1), and was significantly associated with a higher FCS level (mean 37.3 vs 39.5), family APGAR (mean 5.4 vs 7.0), and family well-being (mean 7.1 vs 7.8) controlling for sociodemographic factors (Table 2). Respondents with more family IM chat groups and who received or sent more messages in family IM chat groups were positively associated with a better quality of family communication, family APGAR, and family well-being with dose-response associations (all *P* for trend <.001). More frequent IM chat interactions (received and sent messages combined) was significantly associated with increased FCS, family APGAR, and family well-being levels compared with those of respondents reporting no use. The associations remained robust after additionally adjusting for the frequency of family face-to-face communication (Table 2).

The effect of a greater number of family IM chat groups, received messages, sent messages, and chat interactions (received and sent messages combined) on family functioning and family well-being was largely attenuated after adjusting for FCS (Table 3), providing support for FCS as a mediator. A significant mediation effect of FCS on the above associations (*P* for Sobel test <.001) was found independent of the frequency of family face-to-face communication. Family IM chat interaction accounted for 23.0% of the total effect on family APGAR, and 59.6% of the total effect was mediated by FCS. Similarly, FCS accounted for 51.0% of the 18% total effect of family IM chat interaction on family well-being (*P* for Sobel test <.001).

Table 1. Sociodemographic characteristics according to family instant message chat group use.

Characteristic	No family chat group (n=458), n (%)	Had family chat group (n=1180), n (%)	*P* value	Total (N=1638)	
				Crude, n (%)[a]	Weighted, n (%)[b]
Sex			<.001		
Male	218 (47.6)	423 (35.9)		641 (39.1)	747 (45.6)
Female	240 (52.4)	757 (64.2)		997 (60.9)	891 (54.4)
Age (years)			<.001		
18-24	54 (11.8)	156 (13.2)		210 (12.8)	170 (10.4)
25-44	50 (10.9)	241 (20.4)		291 (17.8)	662 (40.4)
45-64	137 (29.9)	525 (44.5)		662 (40.4)	618 (37.7)
≥65	217 (47.4)	258 (21.9)		475 (29.0)	188 (11.5)
Marital status			.006		
Never been married	115 (25.1)	307 (26.0)		422 (25.7)	293 (17.9)
Married or cohabitating	276 (60.3)	770 (65.3)		1046 (63.9)	1125 (68.7)
Divorced or separated	22 (4.8)	33 (2.8)		55 (3.4)	60 (3.7)
Widowed	45 (9.8)	70 (5.9)		115 (7.0)	160 (9.8)
Education attainment			<.001		
≤Primary	122 (26.6)	123 (10.4)		245 (15.0)	267 (16.3)
Secondary	198 (43.2)	515 (43.6)		713 (43.5)	837 (51.1)
≥Tertiary	138 (30.1)	542 (45.9)		680 (41.5)	534 (32.6)
Monthly household income (HK$)[c]			<.001		
≤9999	128 (28.0)	138 (11.7)		266 (16.2)	170 (10.4)
10,000-19,999	85 (18.6)	145 (12.3)		230 (14.0)	277 (16.9)
20,000-29,999	62 (13.5)	225 (19.1)		287 (17.5)	347 (21.2)
30,000-39,999	41 (8.9)	164 (13.9)		205 (12.5)	241 (14.7)
≥40,000	84 (18.3)	372 (31.5)		456 (27.8)	447 (27.3)
Unstable	58 (12.7)	136 (11.5)		194 (11.8)	156 (9.5)
Frequency of family face-to-face communication (n=1360)			.14		
Often	207 (68.5)	598 (75.1)		959 (70.5)	976 (71.8)
Sometimes	68 (22.5)	149 (18.7)		280 (20.6)	262 (19.3)
Seldom	23 (7.6)	43 (5.4)		93 (6.8)	102 (7.5)
Never	4 (1.3)	6 (0.8)		28 (2.1)	19 (1.4)

[a]Sample size varied because of missing values.

[b]Weighted by the sex-age distributions of the Hong Kong general population in 2015 and the education attainment distribution in 2011.

[c]HK$: Hong Kong dollars; HK $7.8=US $1.

Table 2. Association of the number of family instant message (IM) chat groups (N=1638) and use (N=1180) with family communication, family functioning, and family well-being.

Variable	Respondents, n (%)	Family communication quality (10-50), adjusted β (95% CI)		Family functioning[a] (0-10), adjusted β (95% CI)		Family well-being (0-10), adjusted β (95% CI)	
		Model 1[b]	Model 2[c]	Model 1	Model 2	Model 1	Model 2
Number of family IM groups							
0	458 (28.0)	Reference	Reference	Reference	Reference	Reference	Reference
1	356 (21.7)	.82 (–.76 to 2.42)	.64 (–.88 to 2.15)	.70 (.17 to 1.24)**	.20 (–.55 to .96)	.27 (.04 to .49)*	.11 (–.16 to .38)
2	273 (16.7)	1.02 (–.67 to 2.72)	.86 (–.75 to 2.48)	.52 (–.06 to 1.10)	.19 (–.62 to .99)	.39 (.14 to .64)**	.21 (–.09 to .51)
≥3	551 (33.6)	2.39 (1.01 to 3.76)***	2.04 (.72 to 3.35)**	1.40 (.91 to 1.89)***	1.09 (.43 to 1.75)***	.61 (.40 to .82)***	.39 (.15 to .64)**
P for trend		<.001	.002	<.001	.001	<.001	.001
Number of received IM from family chat groups/day							
<1	94 (8.0)	Reference	Reference	Reference	Reference	Reference	Reference
1-2	338 (28.7)	3.81 (1.21 to 6.40)**	2.73 (.22 to 5.24)*	1.32 (.44 to 2.21)**	.41 (–.84 to 1.66)	.25 (–.10 to .59)	.03 (–.36 to .43)
3-10	533 (45.3)	4.32 (1.85 to 6.78)***	3.35 (.97 to 5.73)**	1.74 (.89 to 2.59)***	.84 (–.35 to 2.03)	.57 (.24 to .90)***	.35 (–.03 to .72)
11-20	110 (9.4)	6.76 (3.53 to 10.01)***	5.39 (2.25 to 8.52)***	2.30 (1.26 to 3.33)***	1.17 (–.39 to 2.72)	.81 (.40 to 1.23)***	.78 (.28 to 1.28)*
>20	102 (8.7)	6.28 (3.23 to 9.32)***	4.96 (2.01 to 7.91)***	2.72 (1.66 to 3.79)***	1.64 (.17 to 3.11)*	.93 (.51 to 1.36)***	.79 (.30 to 1.27)*
P for trend		<.001	<.001	<.001	.007	<.001	<.001
Number of sent IM in family chat groups/day							
<1	192 (16.3)	Reference	Reference	Reference	Reference	Reference	Reference
1-2	503 (42.7)	2.70 (.86 to 4.53)**	2.46 (.70 to 4.22)**	.80 (.17 to 1.43)*	.68 (–.19 to 1.56)	.44 (.19 to .69)***	.39 (.10 to .68)**
3-10	410 (34.8)	3.83 (1.90 to 5.75)***	3.13 (1.27 to 4.99)***	1.35 (.70 to 2.01)***	1.11 (.19 to 2.03)*	.68 (.42 to .94)***	.57 (.27 to .87)***
11-20	49 (4.2)	5.93 (2.61 to 8.19)***	5.32 (2.24 to 8.40)***	1.71 (.57 to 2.86)***	1.11 (–.41 to 2.63)	1.24 (.77 to 1.72)***	1.34 (.79 to 1.89)***
>20	25 (2.0)	4.34 (.11 to 8.57)*	3.50 (–.55 to 7.55)	2.90 (1.33 to 4.47)***	2.96 (.95 to 4.97)**	.78 (.15 to 1.42)**	.62 (–.09 to 1.33)
P for trend		<.001	.001	<.001	.002	<.001	<.001
Frequency of family IM chat interaction[d]		.83 (.47 to 1.19)***	.69 (.34 to 1.04)***	.36 (.23 to .48)***	.27 (.10 to .44)**	.16 (.11 to .21)***	.15 (.09 to .21)***

[a]Family functioning assessed on the Family Adaptability, Partnership, Growth, Affection, and Resolve (APGAR) scale.

[b]Regression model 1: adjusted for sex, age, education attainment, family income, and marital status.

[c]Regression model 2: additionally adjusted for the frequency of face-to-face family communication (often, sometimes, seldom, or never).

[d]Composite variable, frequency of family IM chat interaction (range 0-8), sum of the number of messages received from IM chat groups and number of messages of sent in IM chat groups per day.

*$P<.05$; **$P<.01$; ***$P<.001$.

Table 3. Adjusted indirect, direct, and total effects of the number of family instant message (IM) chat groups (N=1638) and use (N=1180) on family functioning and family well-being mediated by family communication using the Sobel test.

Variable	Number of family IM chat groups	Number of received IM from family chat groups/day	Number of sent IM in family chat groups/day	Frequency of family IM chat interaction[a]
Family functioning				
Total effect, adjusted β[b] (95% CI)	.33 (.13 to .54)**	.34 (.05 to .63)*	.45 (.12 to .78)**	.23 (.06 to .40)**
Indirect effect (through mediator), adjusted β (95% CI)	.14 (.04 to .23)**	.22 (.08 to .36)**	.23 (.09 to .38)**	.14 (.06 to .22)***
Direct effect (without mediator), adjusted β (95% CI)	.20 (–.01 to .40)	.12 (–.16 to .39)	.22 (–.09 to .52)	.09 (–.06 to .24)
Proportion of total effect mediated (%)	40.4	64.9	51.8	59.6
Family well-being				
Total effect, adjusted β (95% CI)	.14 (.03 to .25)*	.31 (.16 to .46)***	.31 (.14 to .49)***	.18 (.10 to .27)***
Indirect effect, adjusted β (95% CI)	.10 (.03 to .16)**	.15 (.06 to .24)***	.16 (.06 to .25)***	.09 (.04 to .14)***
Direct effect, adjusted β (95% CI)	.04 (–.05 to .14)	.16 (.04 to .28)**	.15 (.02 to .29)*	.09 (.02 to .15)**
Proportion of total effect mediated (%)	68.6	48.4	50.6	51.0

[a]Composite variable, frequency of family IM chat interaction (range 0-8), sum of the number of messages received from IM chat groups and number of messages sent in IM chat groups per day.

[b]Regression model 2: adjusted for sex, age, education attainment, family income, marital status, and frequency of family face-to-face communication.

*P<.05; **P<.01; ***P<.001.

Discussion

To our knowledge, this is the first study to report the associations of family IM chat groups with family communication, functioning, and well-being. Family interaction has been strongly associated with family relationships and well-being [5,39]. Our findings are in line with previous findings on the use of IM for family communication predicting greater positive relations and well-being [19]. Three-quarters (73.3%) of the respondents had at least one family IM chat group and almost all (93%) of the family IM chat group users received or sent at least one message daily. Respondents with higher education levels, higher household income, and younger age were more likely to use IM for family communication, which might be explained by their higher accessibility and acceptance of new ICTs [19]. The associations remained robust after adjusting for sociodemographic factors and the frequency of family face-to-face communication. Having at least one family chat group was significantly associated with better family functioning and well-being. A dose-response effect of having more family chat groups was found with better family communication and relationships, in concordance with the Chinese “big family” concept, which values the togetherness, cohesiveness, and harmony of the family over individual expression [40]. The adoption of family groups also responds to a deep cross-cultural need to strengthen and maintain family intimacy, and to establish a coherent identity, especially with geographically separated families [40]. Although the information on which family members were included in the family groups was not collected, our results indicate that setting up family chat groups may increase family cohesion and impose a better family relationship.

The number of messages received (21.9% received over 10 messages daily) exceeded the number of messages sent (8.0% sent over 10 messages daily), likely because messages sent in group chats were received by more than one recipient. Respondents with frequent group interaction might have more concerns over family issues, be more eager to share family life information, and have better management of family relationships, thereby improving family well-being [14]. Prior studies have reported that traditional methods of communication (face-to-face and phone calls) were strongly associated with family well-being [19,22]. Increasing reliance on ICTs in recent years has changed and continues to transform the ways families interact, exchange information, and communicate [41]. The steep growth of IM users and the proliferation of IM app features (eg, video calls, file transfer, news sharing, emoji) have provided the most convenient communication ecosystem for society as a whole [16,29].

The association between family group chat and family well-being was partially mediated by quality communication. Maintaining family well-being is increasingly challenging in

modern societies with busy lifestyles, long work hours (Hong Kong residents work an average of 45 hours a week), and poor work-life balance [42]. People who can utilize effective communication to maintain harmonious marital and family relationships are often healthier and happier than those who are in turbulent family relationships [39]. IM chat adds no time demands for gathering, and may provide opportunity to balance work and life conflicts. Group chats simulate a gathering environment where users can effectively share family information, deal with family problems, and provide support, even from a far distance. Our results, if confirmed by prospective data, could provide solid evidence that the use of family chat groups should be encouraged, especially under a busy society context.

Although specific information topics shared in family groups were not determined in this study, IMs in family chat groups may tend to be tailored to the needs of families rather than to individuals as compared to one-on-one conversations. Family members might be more inclined to share positive family information, express concern and affection, and to respond to and resolve personal or family problems in the chat groups, which are considered the main aspects of family communication quality, and are closely associated with strengthened family bonds and psychological support [39]. Gratitude and happiness expressed during family activities are positively associated with family communication and well-being [43]. Using a synchronous communication platform, IM group chat could facilitate family information exchange [22], thereby contributing to improved family functioning and well-being. Future qualitative studies may consider capturing the detailed content of the group conversation that contributes to family functioning and well-being for guiding quality family communication.

Many studies have focused on adolescents or young adults with respect to the use of IM and well-being owing to their higher tendency of use [44]; however, a huge growing trend among adults and elderly users of IM also requires research to understand the possible individual and family implications [45]. In this sample of Chinese adults, almost half (45%) of older adults (aged 65 or above) had no family chat group. One explanation is their perceived and practical access barriers to new technologies [46]. Lack of cognitive skills, information literacy, and social support in the elderly are also responsible for communication inequalities in a digitally dense society [46,47]. We found no interactions between the younger (aged 18-64 years) and older (aged over 65 years) adults on the association between family IM group chat with family communication and family relationships (see Multimedia Appendix 2), suggesting that communication with family members through chat groups equally improved perceived family relationships in the elderly as in the younger adults. Loneliness, isolation, and suicide attempts are invading the psychological health of the elderly, representing a huge public health concern [48,49], especially for aging societies such as Hong Kong [50]. Involving the elderly to use ICTs to communicate could be practical and effective in maintaining family connection; preventing family and social life isolation; and sustaining a higher level of mental, family, and social well-being.

This study has several limitations. First, landline telephone directories were used to sample the potential respondents, which excluded families with only mobile phones. Our recent landline and online (randomly sampled from a population-representative mobile panel) survey found similar characteristics in terms of sex, age, socioeconomic background, and information seeking [51]. Mobile phone and online surveys are therefore needed to complement the findings. Second, recall bias in reporting the number of family chat groups, and messages received and sent in the groups daily could not be excluded. Third, due to the cross-sectional survey design, we cannot rule out residual confounding or reverse causality. Therefore, prospective studies are needed to further confirm the associations and test for the mediation effect we observed. Indeed, families with better family relationships might tend to use IM group chat more frequently to maintain contact. Nevertheless, the association remained robust after controlling for the frequency of face-to-face communication. Most studies on IM use have focused on teenagers and adolescents owing to their higher acceptance of new technologies and lower self-control to behavioral addiction. We used simple measurements for the use of family chat groups, and the content of messages was not assessed, which warrants future investigation. We only investigated the adult population, who have more concerns and tend to pay much more attention to family relationships; thus, whether the family IM group chat equally benefits the younger generation remains unclear. Family health, harmony, and happiness have been identified as determinants of family well-being in Chinese culture [3,25]; however, the perceptions of family well-being might be different in other countries, which affects the generalizability of our findings. The validity of family well-being was indirectly supported by the consistent results with the FCS and family functioning scale, which are validated scales frequently used in Western countries.

In summary, family IM chat use was associated with higher family functioning and well-being, and the association was partially mediated by family communication.

Acknowledgments

The project was funded by the Hong Kong Jockey Club Charities Trust. We would like to thank all the participants who participated in the telephone surveys and the POP (University of Hong Kong) for conducting the surveys.

Authors' Contributions

MW, SC, and TL conceived of the study. SZ and MW analyzed the data. SZ, MW, AL, and TL interpreted the data. SZ wrote the first draft of the manuscript. All authors critically revised and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Means and associations of the number of family instant message chat groups (N=1638) and use (N=1180) with family communication, family functioning, and family well-being.
[DOCX File , 18 KB - jmir_v23i3e18876_app1.docx]

Multimedia Appendix 2
Moderating effects of age on the association of the number of family instant message chat groups (N=1638) and use (N=1180) with family functioning and well-being.
[DOCX File , 19 KB - jmir_v23i3e18876_app2.docx]

References

1. Lewandowski AS, Palermo TM, Stinson J, Handley S, Chambers CT. Systematic review of family functioning in families of children and adolescents with chronic pain. J Pain 2010 Nov;11(11):1027-1038 [FREE Full text] [doi: 10.1016/j.jpain.2010.04.005] [Medline: 21055709]
2. García-Huidobro D, Puschel K, Soto G. Family functioning style and health: opportunities for health prevention in primary care. Br J Gen Pract 2012 Mar;62(596):e198-e203 [FREE Full text] [doi: 10.3399/bjgp12X630098] [Medline: 22429437]
3. Lam WWT, Fielding R, McDowell I, Johnston J, Chan S, Leung GM, et al. Perspectives on family health, happiness and harmony (3H) among Hong Kong Chinese people: a qualitative study. Health Educ Res 2012 Oct;27(5):767-779. [doi: 10.1093/her/cys087] [Medline: 22907531]
4. Zubrick SR, Williams AA, Silburn SR, Vimpani G. Indicators of Social and Family Functioning. Australia: Department of Family and Community Services; May 2000.
5. Segrin C. Family interactions and well-Being: integrative perspectives. J Family Commun 2006 Jan;6(1):3-21. [doi: 10.1207/s15327698jfc0601_2]
6. Albada K, Moore J. Moving from positive thinking to positive talk: Implications for relational well-being. In: Pitts MJ, Soha TJ, editors. Health Communication Vol. 3: Positive Communication in Health and Wellness. New York, NY: Peter Lang Publishing; 2013:117.
7. Segrin C. Interpersonal processes in psychological problems. New York, NY: Guilford Press; 2001.
8. Luk TT, Wang MP, Leung LT, Chen J, Wu Y, Lam TH, et al. Perceived family relationship quality and use of poly-tobacco products during early and late adolescence. Addict Behav 2018 Oct;85:38-42. [doi: 10.1016/j.addbeh.2018.05.011] [Medline: 29843039]
9. Stern MJ, Messer C. How family members stay in touch: a quantitative investigation of core family networks. Marriage Fam Rev 2009 Oct 21;45(6-8):654-676. [doi: 10.1080/01494920903224236]
10. Nardi B, Whittaker S, Bradner E. Interaction and outeraction: instant messaging in action. : ACM Press; 2000 Presented at: ACM Conference on Computer Supported Cooperative Work (CSCW 2000); 2000; Philadelphia, PA. [doi: 10.1145/358916.358975]
11. Elhai JD, Dvorak RD, Levine JC, Hall BJ. Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology. J Affect Disord 2017 Jan 01;207:251-259. [doi: 10.1016/j.jad.2016.08.030] [Medline: 27736736]
12. Elhai JD, Levine JC, Dvorak RD, Hall BJ. Non-social features of smartphone use are most related to depression, anxiety and problematic smartphone use. Comput Hum Behav 2017 Apr;69:75-82. [doi: 10.1016/j.chb.2016.12.023]
13. Horwood S, Anglim J. Problematic smartphone usage and subjective and psychological well-being. Comput Hum Behav 2019 Aug;97:44-50. [doi: 10.1016/j.chb.2019.02.028]
14. Carvalho J, Francisco R, Relvas AP. Family functioning and information and communication technologies: How do they relate? A literature review. Comput Hum Behav 2015 Apr;45:99-108. [doi: 10.1016/j.chb.2014.11.037]
15. Lapierre MA, Lewis MN. Should it stay or should it go now? Smartphones and relational health. Psych Pop Media Cult 2018 Jul;7(3):384-398. [doi: 10.1037/ppm0000119]
16. Lenhart A, Rainie L, Lewis O, Pew Internet and American Life Project. Teenage life online: The rise of the instant-message generation and the Internet's impact on friendships and family relationships. Washington, DC: Pew Internet and American Life Project; Jun 2001.
17. Baym NK, Zhang YB, Lin M. Social interactions across media. New Media Soc 2016 Jun 30;6(3):299-318. [doi: 10.1177/1461444804041438]
18. Ramirez A, Broneck K. 'IM me': Instant messaging as relational maintenance and everyday communication. J Soc Pers Relation 2009 Jul 31;26(2-3):291-314. [doi: 10.1177/0265407509106719]

19. Wang MP, Chu JTW, Viswanath K, Wan A, Lam TH, Chan SS. Using information and communication technologies for family communication and its association with family well-being in Hong Kong: FAMILY Project. J Med Internet Res 2015 Aug 24;17(8):e207 [FREE Full text] [doi: 10.2196/jmir.4722] [Medline: 26303434]
20. Chan M. Mobile phones and the good life: Examining the relationships among mobile use, social capital and subjective well-being. New Media Soc 2013 Dec 16;17(1):96-113. [doi: 10.1177/1461444813516836]
21. Herbsleb J, Atkins D, Boyer D, Handel M, Finholt T. Introducing instant messaging and chat in the workplace. 2002 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 2002; Minneapolis, MN p. 171-178. [doi: 10.1145/503376.503408]
22. Shen C, Wang MP, Chu JT, Wan A, Viswanath K, Chan SSC, et al. Sharing family life information through video calls and other information and communication technologies and the association with family well-being: population-based survey. JMIR Ment Health 2017 Nov 23;4(4):e57 [FREE Full text] [doi: 10.2196/mental.8139] [Medline: 29170145]
23. Shen C, Wang MP, Wan A, Viswanath K, Chan SSC, Lam TH. Health information exposure from information and communication technologies and its associations with health behaviors: Population-based survey. Prev Med 2018 Aug;113:140-146. [doi: 10.1016/j.ypmed.2018.05.018] [Medline: 29787775]
24. Karakas F, Dean Lee M, MacDermid SM. A qualitative investigation into the meaning of family well‐being from the perspective of part‐time professionals. Equal Opportunities Int 2004 Jan;23(1/2):57-77. [doi: 10.1108/02610150410787855]
25. Chan SSC, Viswanath K, Au DWH, Ma CMS, Lam WWT, Fielding R, et al. Hong Kong Chinese community leaders' perspectives on family health, happiness and harmony: a qualitative study. Health Educ Res 2011 Aug;26(4):664-674. [doi: 10.1093/her/cyr026] [Medline: 21536713]
26. Thematic Household Survey Report No 44: Relationships among Family Members. Hong Kong: Census and Statistics Department URL: https://www.statistics.gov.hk/pub/B11302442010XXXXB0100.pdf [accessed 2020-11-25]
27. Thematic Household Survey Report No 67: Information technology usagepenetration. Hong Kong: Census and Statistics Department URL: https://www.statistics.gov.hk/pub/B11302672019XXXXB0100.pdf [accessed 2020-11-25]
28. Bano S, Cisheng W, Khan AN, Khan NA. WhatsApp use and student's psychological well-being: Role of social capital and social integration. Child Youth Serv Rev 2019 Aug;103:200-208. [doi: 10.1016/j.childyouth.2019.06.002]
29. Flanagin AJ. IM online: instant messaging use among college students. Commun Res Rep 2005 Aug;22(3):175-187. [doi: 10.1080/00036810500206966]
30. Wang MP, Wang X, Viswanath K, Wan A, Lam TH, Chan SS. Digital inequalities of family life information seeking and family well-being among Chinese adults in Hong Kong: a population survey. J Med Internet Res 2014 Oct 03;16(10):e227 [FREE Full text] [doi: 10.2196/jmir.3386] [Medline: 25281831]
31. Guo N, Wang MP, Luk TT, Ho SY, Fong DYT, Chan SS, et al. The association of problematic smartphone use with family well-being mediated by family communication in Chinese adults: A population-based study. J Behav Addict 2019 Sep 01;8(3):412-419 [FREE Full text] [doi: 10.1556/2006.8.2019.39] [Medline: 31394912]
32. Hong Kong as an Information Society. Census and Statistics Department. Hong Kong: The Government of the Hong Kong Special Administrative Region URL: https://www.censtatd.gov.hk/hkstat/sub/sp120.jsp?productCode=B1110006 [accessed 2020-11-25] [WebCite Cache ID https://www.censtatd.gov.hk/hkstat/sub/sp120.jsp?productCode=B1110006]
33. Olson D. FACES IV and the Circumplex Model: validation study. J Marital Fam Ther 2011 Jan;37(1):64-80. [doi: 10.1111/j.1752-0606.2009.00175.x] [Medline: 21198689]
34. Smilkstein G, Ashworth C, Montano D. Validity and reliability of the family APGAR as a test of family function. J Fam Pract 1982 Aug;15(2):303-311. [Medline: 7097168]
35. Wu CST, Wong HT, Yu KF, Fok KW, Yeung SM, Lam CH, et al. Parenting approaches, family functionality, and internet addiction among Hong Kong adolescents. BMC Pediatr 2016 Aug 18;16:130 [FREE Full text] [doi: 10.1186/s12887-016-0666-y] [Medline: 27538688]
36. Chan SSC, Viswanath K, Au DWH, Ma CMS, Lam WWT, Fielding R, et al. Hong Kong Chinese community leaders' perspectives on family health, happiness and harmony: a qualitative study. Health Educ Res 2011 Aug;26(4):664-674. [doi: 10.1093/her/cyr026] [Medline: 21536713]
37. Shen C, Wang MP, Ho HCY, Wan A, Stewart SM, Viswanath K, et al. Test-retest reliability and validity of a single-item Self-reported Family Happiness Scale in Hong Kong Chinese: findings from Hong Kong Jockey Club FAMILY Project. Qual Life Res 2019 Feb;28(2):535-543. [doi: 10.1007/s11136-018-2019-9] [Medline: 30284182]
38. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986 Dec;51(6):1173-1182. [doi: 10.1037//0022-3514.51.6.1173] [Medline: 3806354]
39. Sillars A, Canary D, Tafoya M. Communication, conflict, and the quality of family relationships. In: Vangelisti AL, editor. LEA's communication series. Handbook of family communication. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2004:413-446.
40. Chang H. Harmony as Performance: The Turbulence Under Chinese Interpersonal Communication. Discourse Stud 2001 May;3(2):155-179. [doi: 10.1177/1461445601003002001]
41. Galvin K, Braithwaite D, Bylund C. Family Communication: Cohesion and Change, 10th edition. UK: Taylor and Francis; 2018.

42. Report on Annual Earnings and Hours Survey. Census and Statistics Department.: The Government of Hong Kong and Special Administrative Region URL: https://www.censtatd.gov.hk/hkstat/sub/sp210.jsp?productCode=B1050014 [accessed 2020-11-25]
43. Ho HCY, Mui M, Wan A, Ng Y, Stewart SM, Yew C, et al. Happy family kitchen II: a cluster randomized controlled trial of a community-based family intervention for enhancing family communication and well-being in Hong Kong. Front Psychol 2016;7:638. [doi: 10.3389/fpsyg.2016.00638] [Medline: 27199864]
44. Sohn S, Rees P, Wildridge B, Kalk NJ, Carter B. Prevalence of problematic smartphone usage and associated mental health outcomes amongst children and young people: a systematic review, meta-analysis and GRADE of the evidence. BMC Psychiatry 2019 Nov 29;19(1):356 [FREE Full text] [doi: 10.1186/s12888-019-2350-x] [Medline: 31779637]
45. Public health implications of excessive use of the Internet, computers, smartphones and similar electronic devices. World Health Organization. 2014. URL: https://www.who.int/news/item/13-09-2018-public-health-implications-of-excessive-use-of-the-internet-and-other-communication-and-gaming-platforms [accessed 2021-02-25]
46. Porter CE, Donthu N. Using the technology acceptance model to explain how attitudes determine Internet usage: The role of perceived access barriers and demographics. J Bus Res 2006 Sep;59(9):999-1007. [doi: 10.1016/j.jbusres.2006.06.003]
47. Viswanath K, Ackerson LK. Race, ethnicity, language, social class, and health communication inequalities: a nationally-representative cross-sectional study. PLoS One 2011 Jan 18;6(1):e14550 [FREE Full text] [doi: 10.1371/journal.pone.0014550] [Medline: 21267450]
48. Singh A, Misra N. Loneliness, depression and sociability in old age. Ind Psychiatry J 2009 Jan;18(1):51-55 [FREE Full text] [doi: 10.4103/0972-6748.57861] [Medline: 21234164]
49. Frierson RL. Suicide attempts by the old and the very old. Arch Intern Med 1991 Jan 01;151(1):141. [doi: 10.1001/archinte.1991.00400010143022]
50. Chi I, Chou KL. Social support and depression among elderly Chinese people in Hong Kong. Int J Aging Hum Dev 2001;52(3):231-252. [doi: 10.2190/V5K8-CNMG-G2UP-37QV] [Medline: 11407488]
51. Zhao SZ, Wong JYH, Wu Y, Choi EPH, Wang MP, Lam TH. Social distancing compliance under COVID-19 pandemic and mental health impacts: a population-based study. Int J Environ Res Public Health 2020 Sep 14;17(18):6692 [FREE Full text] [doi: 10.3390/ijerph17186692] [Medline: 32937929]

Abbreviations

3Hs: perceived family harmony, happiness, and health
APGAR: Adaptation, Partnership, Growth, Affection, and Resolve
FCS: Family Communication Scale
FHInTS: Family and Health Information Trends Survey
HK $: Hong Kong dollars
ICT: Information and communication technologies
IM: instant messaging
POP: Public Opinion Programme
SES: socioeconomic status

Edited by G Eysenbach; submitted 25.03.20; peer-reviewed by J Sin, E Ennis; comments to author 29.06.20; revised version received 02.08.20; accepted 23.11.20; published 15.03.21.

Please cite as:
Zhao SZ, Luk TT, Guo N, Wang MP, Lai AYK, Wong BYM, Fong DYT, Chan SSC, Lam TH
Association of Mobile Instant Messaging Chat Group Participation With Family Functioning and Well-Being: Population-Based Cross-sectional Study
J Med Internet Res 2021;23(3):e18876
URL: https://www.jmir.org/2021/3/e18876
doi:10.2196/18876
PMID:33720034

Original Paper

Trends and Characteristics of the US Adult Population's Behavioral Patterns in Web-Based Prescription Filling: National Survey Study

Lin-Ya Yang[1], MSc; Jennifer G Lyons[2], PhD; Steven R Erickson[3], PharmD; Chung-Hsuen Wu[1], PhD

[1]School of Pharmacy, College of Pharmacy, Taipei Medical University, Taipei, Taiwan

[2]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, United States

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

Corresponding Author:
Chung-Hsuen Wu, PhD
School of Pharmacy
College of Pharmacy
Taipei Medical University
No 250, Wu-hsing St
Xinyi District
Taipei, 11031
Taiwan
Phone: 886 227361661 ext 6180
Fax: 886 227390671
Email: chunghwu@tmu.edu.tw

Abstract

Background: Filling a prescription on the web has become an alternative to in-person pharmacies for individuals to access their medications. However, the adoption of web-based filling has been gradual, and the use patterns remain to be unclear.

Objective: This study aims to estimate the trend and prevalence of web-based prescription-filling behavior and identify associated factors among adults in the United States.

Methods: We used data from the US National Health Interview Survey (NHIS) from 2009 to 2018. Adult respondents (aged ≥18 years and over) self-reported their behavior of web-based prescription filling, which was defined as having filled a prescription using the internet in the past 12 months during the survey year. We reported trends using weighted percentages adjusted by the NHIS complex sampling design. We used descriptive statistics and multivariable logistic regression models to examine trends and identify factors associated with web-based prescription-filling behavior.

Results: The estimated number of adults reporting web-based prescription-filling behavior significantly increased from 13,319,877 (13,319,877/225,217,942, 5.91%) in 2009 to 28,308,262 (28,308,262/246,611,125, 11.48%) in 2018 *(P*<.001). Those who were more likely to report filling a prescription on the web were aged between 35 and 74 years, female, White, and frequent users of the computer or internet; these adults also reported higher education, higher income, insurance coverage, and poorer health status.

Conclusions: Web-based prescription-filling behavior among US adults has increased significantly from 2009 to 2018. Health care providers should be aware of the upward trend in the use of web-based pharmacies and ensure the clinical safety of web-based prescriptions.

(J Med Internet Res 2021;23(3):e23662) doi:10.2196/23662

KEYWORDS

internet; prescription; online pharmacy; National Health Interview Survey; trend

Introduction

Health Information and Internet

The widespread use of the internet as a means of obtaining health information and health care has a tremendous impact on health care systems [1,2]. It is estimated that more than 70% of internet users globally search for health information on the web [3-5]. The prevalence of web-based pharmacy use has increased over time [6,7]. Web-based pharmacies, which include web-based ordering and mail delivery, have become an alternative to obtaining prescription medications from in-person pharmacies. For example, more than 14% of internet users in

the United States reported that they had ever bought medicines on the web [8]. According to data from the Health Information National Trends Survey (HINTS), the percentage of people in the United States who purchased medications on the web increased from 9.1% in 2003 to 20.2% in 2013 [7].

Benefits and Concerns

Individuals may be increasingly interested in purchasing prescription medications on the web for several reasons. First, purchasing medication on the web is more convenient than an in-person pharmacy because web-based pharmacies are available 24 hours a day. People who have difficulty traveling to a physical pharmacy can more easily order medications on the web from home [9,10]. Second, purchasing on the web is more private and attractive for individuals who are concerned with obtaining their medications publicly from a community pharmacy [9,10]. Finally, given that the cost of prescription medication continues to increase in the United States [11], saving money is the leading reason for purchasing medication on the internet, although mostly from foreign websites [12,13].

Despite the perceived benefits, there remain several concerns regarding the use of web-based pharmacies to obtain prescriptions. Underregulated or illegal web-based pharmacies can sell medications that may contain the wrong dosage or active ingredients or be past the expiration date or are not approved by the US Food and Drug Administration [10,14]. The 2018 report of the National Association of Boards of Pharmacy (NABP) showed that among the 11,943 web-based pharmacies reviewed, the vast majority (95%) were found to be noncompliant with relevant federal or state pharmacy laws [15]. Furthermore, less than 3.0% of the reviewed web-based pharmacies were accredited by the NABP-verified internet pharmacy practice sites [15].

Purchasing Medicines on the Web

Several previous studies have examined web-based purchasing medications [1,8,16-22]. These studies have explored the characteristics of purchasers using web-based pharmacies [1,8,16,18-22], the experience of purchasing medications on the web [16,17,19-21], reasons for purchasing medications on the web [19,20,22], and patients' tendency to disclose to their health care providers that they purchase medications using web-based pharmacies [8]. However, most of the previous research has studied web-based medication-purchasing behaviors from other countries [16,17,19,20,22] but not specifically from the United States. Moreover, one difference between the United States and many other countries is that prescription-only medication can be purchased using web-based pharmacies that are located within the United States [16,17,19,20,22]. Rules vary by country, from very restrictive (no internet pharmacy allowed) to allowing the purchase of prescription medication from another country. For example, unlike the United States, some other countries allow only nonprescription medications to be purchased on the web [23]. Finally, the data obtained are typically from specific hospitals [17,21] or community pharmacies [16] or from mailed surveys of a specific sample [20,22], not using a nationally representative survey design. Only a few studies used nationally representative US data to investigate web-based pharmacy use [1,8,18]. For example, one study was conducted using the Medical Expenditure Panel Survey (MEPS) to evaluate characteristics associated with web-based pharmacy use [18], whereas another study was conducted using the HINTS to evaluate the characteristics of people purchasing medication on the web [8]. However, the findings from the MEPS and HINTS studies were published 10 years ago, and more updated evidence is needed to understand newer trends in the use of web-based pharmacies.

Objective

With the expansion of internet use [24] and the recent COVID-19 pandemic [25], purchasing medications using web-based pharmacies may be the new normal for individuals to access their medications in the future. Therefore, it is important to update the information available regarding the prevalence and utilization patterns of people filling prescriptions using web-based pharmacies. Using data from the National Health Interview Survey (NHIS), which is a nationally representative data set of the United States, provides this unique opportunity. This study aims to (1) evaluate the trend and (2) identify the characteristics associated with web-based prescription-filling behavior among the US adult population.

Methods

Data Source

We used data from the US NHIS, from 2009 to 2018, which is a cross-sectional household interview survey designed to track health status, health care access, and health resource utilization among the civilian noninstitutionalized population in the United States. [26]. The NHIS was initiated in 1957 and is conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), which releases the survey data annually [26].

The data in the NHIS were obtained through a complex sample design involving stratification, clustering, and weighting. Given the changes in the distribution of the US population, a new sample design was implemented in 2016 [27]. The data of the NHIS are collected from face-to-face interviews. Therefore, to keep the data collection more manageable and cost-effective, the sample was systematically selected using multistage sampling technologies (survey data before 2016 NHIS) [28] and geographically clustered sampling techniques (survey data from 2016 NHIS) [27]. Multistage sampling technologies partition the US population into several nested levels of strata and clusters consisting of a sample of primary sampling units (PSUs) [28]. Respondents were selected from the PSUs to form the sample. Regarding the geographically clustered sampling techniques, the sampling process started by dividing the United States into many geographic areas. One geographic area included 1 or 2 strata, and 1 stratum contained many clusters of addresses. Then, a specific number of clusters according to the size of the strata would be selected for the NHIS sample [26,27].

We used data from the Household File, Family File, Person File, and Sample Adult File from the NHIS for this study. From 2009 to 2018, the survey contained approximately 35,000 households, yielding about 87,500 respondents who participated in the survey each year [26]. The number of respondents can

be weighted to obtain estimates of the US population. The response rates each year were 82.2%, 79.5%, 82.0%, 77.6%, 75.7%, 73.8%, 70.1%, 67.9%, 66.5%, and 64.2% [26].

In 2009, the NHIS started collecting data on the use of health information technology in the Sample Adult File. A total of 5-10 questions, depending on the year, about using the internet to search for health information, learn about health topics, fill a prescription, schedule medical appointments, and email health care providers were asked in a face-to-face interview [29]. The NHIS did not release these data in 2010. More detailed information about study design, interview procedure, data editing, and survey questionnaires is available on the official website.

Study Population

All respondents aged ≥18 years who participated in the survey were included as the study population.

Study Outcome (Web-Based Prescription-Filling Behavior)

The survey contained one question asking respondents if they had ever filled a prescription on the internet during the past 12 months. In 2009, the question was formatted as, "Did you refill a prescription on the internet in the past 12 months?" In the 2011 to 2018 surveys, the question was formatted as "DURING THE PAST 12 MONTHS, have you ever used computers for any of the following … Fill a prescription." Respondents who answered "yes" to the question were defined as fillers of web-based prescription filling, and those who answered "no" were enrolled as nonfillers of web-based prescription filling.

Covariates (Factors and Characteristics)

We used the Anderson Behavioral Model of Health Services Use as a conceptual framework to identify factors associated with the behavior of web-based prescription filling [30-32]. The Anderson model is widely used to study health behaviors and health service utilization [30-32]. According to the model, factors that are associated with health behavior or health services utilization can be classified into predisposing, enabling, and need factors. In this study, we identified all the variables from the Family Files, Person Files, and Sample Adult Files in the NHIS and further divided them into 3 categories based on the Anderson model.

Predisposing factors included age (range: 18-34, 35-49, 50-64, 65-74, and ≥75 years), gender (male and female), race (non-Hispanic White, Hispanic, non-Hispanic Black, non-Hispanic Asian, and non-Hispanic others), education (less than high school, high school, college, and higher than college), region (Northwest, Midwest, South, and West), marital status (married or living with partner, divorced or separated or widowed, and single or never married), work status (had job and no job), frequency of computer usage (never, some days, most days, and every day), and frequency of internet usage (never, per day, per week, per month, and per year). Enabling factors included family income (<US $35,000, US $35,000-49,999, US $45,000-74,999, US $75,000-$99,999, and ≥US $100,000) and health insurance coverage (covered and not covered). Needs factors included perceived health status (excellent, very good, good, fair, and poor), mobility limitations (yes or no), prescribed medication by doctors or health professionals (yes or no), and drugs bought from another country to save money (yes or no). It should be noted that some variables, including frequency of computer usage, frequency of internet usage, prescribed medication by doctors or health professionals, and drugs bought from another country to save money, were only recorded in the 2018 survey but not in the 2009 survey. Therefore, these variables can only be analyzed using the 2018 survey data.

Statistical Analysis

A total of 3 major statistical analyses were performed in our study. First, we estimated the annual weighted percentage of web-based prescription-filling behavior among US adults from 2009 to 2018. To evaluate the trend change in web-based prescription-filling behavior, we compared the prevalence between 2009 and 2018. In addition to the overall trend of web-based prescription-filling behavior, we also examined the trends stratified by demographic subpopulation, including age, gender, race, and education levels. Second, we compared selected factors among respondents in 2009 and 2018 separately, using descriptive statistics to identify changes in patterns of web-based prescription-filling behavior. Wald chi-square tests were used to compare the categorized characteristics between web-based prescription fillers and nonfillers. Finally, we conducted multivariable logistic regression to evaluate the associations between factors and web-based prescription-filling behavior among the study population in 2018.

The NHIS uses a complex sample design (stratification, clustering, and weighting) to represent the US population. Sampling weights are necessary for numbers from respondents to produce representative national estimates [26]. The weight is the person-level base weight, which is adjusted according to poststratification and calibration by age, sex, race, and ethnicity classes based on the 2000 US census for survey in 2009 to 2015 and the 2010 census for survey in 2016 to 2018 [26,28].

All statistical procedures were performed using SAS software (version 9.4; SAS Institute). The SAS survey procedure (proc surveymean, proc surveyfreq, and proc surveylogistic) was used to perform the analyses after considering the stratification, clustering, and weighting of the NHIS sample design to represent the US population and prevent underestimation of variances [33]. The two-sided test was used, and statistical significance was set at $P<.05$. As the NHIS data are publicly available and deidentified, this study was exempt from a full ethical review by the Joint Institutional Review Board of Taipei Medical University.

Results

Trend and Prevalence of Web-Based Prescription Fill

Figure 1 shows the trend of web-based prescription fill from 2009 to 2018 among adults in the United States. All the numbers were weighted to obtain US population estimates. The estimated number of respondents in the United States filling the prescription on the web significantly grew from 13.3 million in 2009 to 28.3 million in 2018 ($P<.001$). The overall annual

prevalence of web-based prescription filling was relatively constant between 2009 (13,319,877/225,217,942, 5.91%) and 2014 (16,072,655/236,126,315, 6.81%) but increased from 2014 (16,072,655/236,126,315, 6.81%) to 2018 (28,308,262/246,611,125, 11.48%).

The trends for demographic subpopulations are shown in Figure 2. In subpopulations, trends in the prevalence of web-based prescription-filling behavior were similar to the overall trend in the general adult population. In the trend of web-based prescription-filling behavior by age, respondents aged between 18 and 24 years and ≥75 years were the least likely to fill a prescription on the internet. In contrast, respondents aged 35-74 years were more likely to fill prescriptions on the web, and the percentage significantly increased during the past decade (P<.001).

In terms of gender differences, the prevalence of web-based prescription-filling behavior was consistently greater in women than in men from 2009 to 2018. In both genders, the behavior of web-based prescription filling significantly increased from 2009 to 2018 (P<.001). Among all ethnic groups, non-Hispanic White respondents were most likely to fill the prescription on the web, followed by non-Hispanic Asian respondents (P<.001). The higher the education level, the more likely adults would choose to fill a prescription on the web (P<.001). The prevalence of web-based prescription-filling behavior for respondents with a higher than college degree was consistently the highest among adult respondents in the United States. In contrast, adults with an education level less than high school were less likely to fill a prescription on the web, and the percentage of use remained consistently low from 2009 to 2018.

Figure 1. Trends in web-based prescription-filling behavior among US adults from 2009 to 2018. The prevalence of online prescription fill among adults in the US was found to significantly increase from 2009 (N=13,319,877) to 2018 (N=28,308,262; P<.01). The National Health Interview Survey did not release the data related to online prescription fill in the 2010 survey.

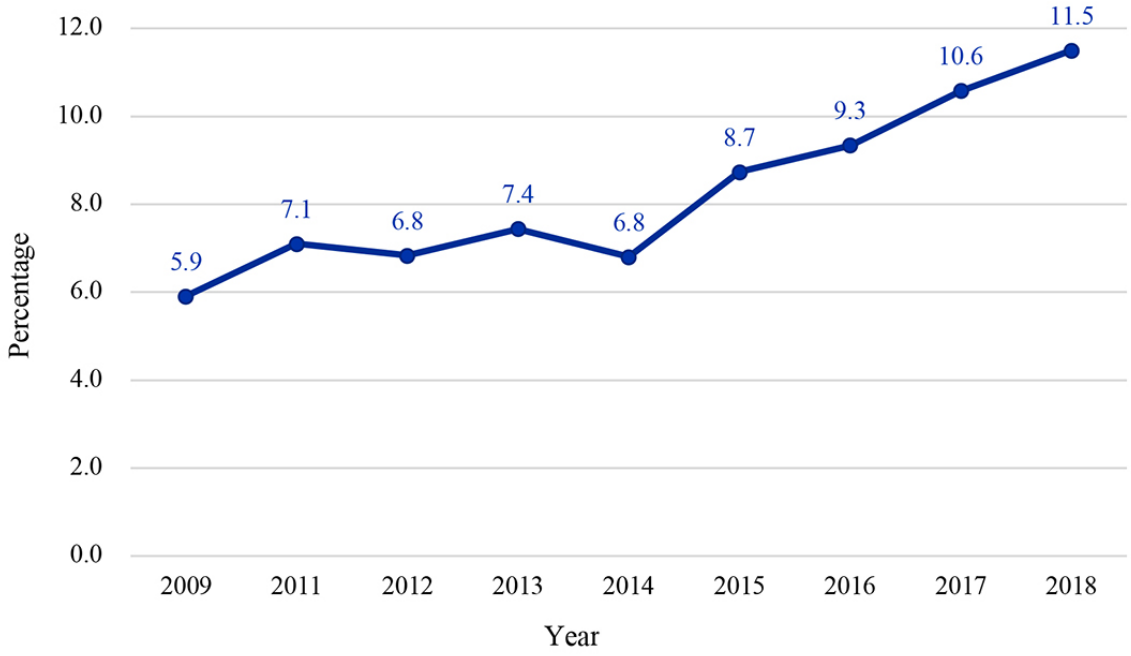
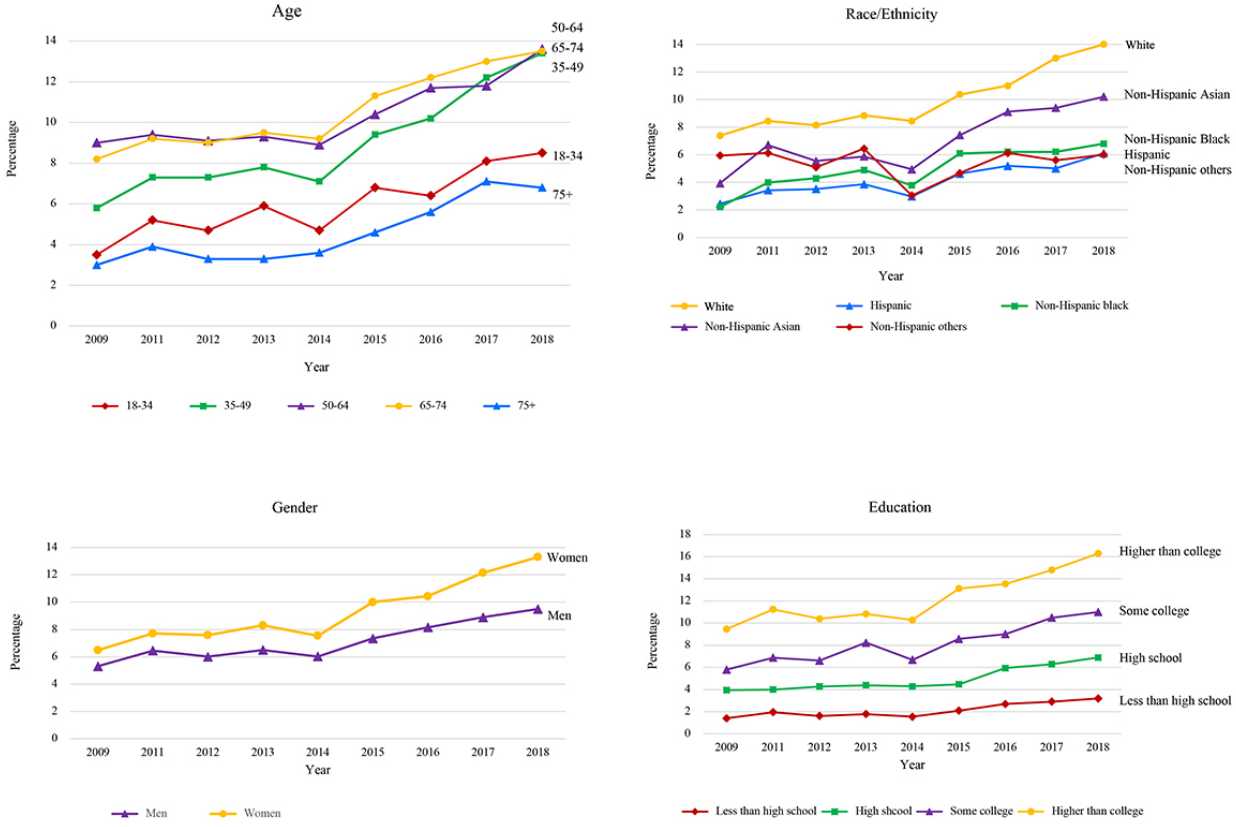


Figure 2. Prevalence of web-based prescription-filling behaviors among adults in the United States by demographic variables, 2009 to 2018.



Characteristics of Study Populations

Table 1 shows the characteristics of the 2009 and 2018 study populations. In both 2009 and 2018, several characteristics, such as age (*P*<.001), sex (*P*=.003 in 2009 and *P*<.001 in 2018), race or ethnicity (*P*<.001), education (*P*<.001), marital status (*P*<.001), work status (*P*<.001), family income (*P*<.001), and health insurance coverage (*P*<.001), were associated with a higher probability of web-based prescription-filling behavior. Moreover, variables including frequency of computer usage (*P*<.001), frequency of internet usage (*P*<.001), prescribed medication by doctors or health professionals (*P*<.001), and having bought drugs from another country to save money (*P*<.001) were found to be significantly associated with web-based prescription-filling behavior in 2018.

Within each characteristic, respondents aged between 50 and 64 years, female, White, and who had higher education, higher income, and insurance coverage were significantly more likely to fill prescriptions on the web in both 2009 and 2018. In addition, respondents who were married or lived with a partner were significantly more likely to fill a prescription on the web when compared with those who were divorced or widowed or single or never married (10,012,769/138,405,448, 7.23% vs 1,659,300/38,520,573, 4.31% vs 1,632,904/48,049,055, 3.40%; *P*<.001 in 2009; 19,956,831/147,945,646, 13.50% vs 3,757,131/40,811,170, 9.21% and 4,584,495/57,461,184, 8.00%; *P*<.001 in 2018).

In terms of working status, respondents who had a job in the past 12 months were more likely to complete the prescription on the web than respondents without a job (10,172,125/157,802,261, 6.4% vs 3,147,752/67,277,067, 4.7%; *P*<.001 in 2009; 20,794,336/169,686,451, 12.3% vs 7,486,229/76,519,060, 9.8%; *P*<.001 in 2018). Perceived health status and region were significantly associated with web-based prescription filling in 2018 (*P*<.001 and *P*<.001, respectively) but not in 2009. In 2018, the region was significantly different, with people in the Midwest and West having a higher filling behavior compared with people in the Northwest or the South.

Table 1. Characteristics and a comparison between web-based prescription fillers and nonfillers in the past 12 months during the survey year: weighted statistics in 2009 (n=1489; N=13,319,877)[a] and 2018 (n=2892; N=28,308,262)[a].

Variables	2009			2018		
	Yes	No	*P* value	Yes	No	*P* value
Predisposing factors, weighted n (%)						
Age (years)			<.001			<.001
18-34	2,424,539 (3.52)	66,471,375 (96.48)		6,200,449 (8.49)	66,865,684 (91.51)	
35-49	3,643,127 (5.80)	59,175,765 (94.20)		8,149,953 (13.44)	52,483,963 (86.56)	
50-64	5,070,551 (9.03)	51,100,111 (90.97)		8,440,377 (13.62)	53,550,362 (86.38)	
65-74	1,668,375 (8.18)	18,729,467 (91.82)		4,126,912 (13.53)	26,381,215 (86.47)	
75+	513,285 (3.03)	16,421,347 (96.97)		1,390,571 (6.81)	19,021,639 (93.19)	
Gender			.003			<.001
Male	5,771,128 (5.31)	102,970,784 (94.69)		11,324,986 (9.51)	107,766,068 (90.49)	
Female	7,548,749 (6.48)	108,927,281 (93.52)		16,983,276 (13.32)	110,536,795 (86.68)	
Race			<.001			<.001
Non-Hispanic White	11,436,419 (7.38)	143,504,508 (92.62)		22,072,419 (14.01)	135,502,059 (85.99)	
Hispanic	759,766 (2.44)	30,349,873 (97.56)		2,444,137 (6.05)	37,945,521 (93.95)	
Non-Hispanic Black	589,266 (2.21)	26,039,713 (97.79)		2,033,915 (6.76)	28,056,833 (93.24)	
Non-Hispanic Asian	411,558 (3.93)	10,057,804 (96.07)		1,565,100 (10.20)	13,773,476 (89.90)	
Non-Hispanic others	122,868 (5.94)	1,946,167 (94.06)		192,691 (5.99)	3,024,974 (84.01)	
Education			<.001			<.001
Less than high school	563,916 (1.41)	39,347,807 (98.59)		1,098,243 (3.19)	33,369,627 (96.81)	
High school	2,189,184 (3.95)	53,292,533 (96.05)		3,721,460 (6.94)	49,863,572 (93.06)	
College	2,621,675 (5.80)	42,561,073 (94.20)		5,036,039 (10.96)	40,928,096 (89.04)	
Higher than college	7,407,794 (9.47)	70,847,409 (90.53)		16,952,950 (16.35)	86,765,573 (83.65)	
Region			.43			<.001
Northwest	2,200,012 (5.61)	37,027,777 (94.39)		4,128,255 (9.69)	38,483,459 (90.31)	
Midwest	3,250,003 (5.96)	51,305,184 (94.04)		7,115,471 (13.10)	47,180,604 (86.90)	
South	4,566,581 (5.67)	76,000,333 (94.33)		9,709,220 (10.67)	81,307,356 (89.33)	
West	3,303,281 (6.49)	47,564,771 (93.51)		7,355,316 (12.53)	51,331,444 (87.47)	
Marital status			<.001			<.001
Married or living with partner	10,012,769 (7.23)	128,392,679 (92.77)		19,956,831 (13.49)	127,988,815 (86.51)	
Divorced, separated, widowed	1,659,300 (4.31)	36,861,273 (95.69)		3,757,131 (9.21)	37,054,039 (90.79)	
Single, never married	1,632,904 (3.40)	46,416,151 (96.60)		4,584,495 (7.98)	52,876,689 (92.02)	
Work status (past 12 months)			<.001			<.001
Had job	10,172,125 (6.45)	147,630,136 (93.55)		20,794,336 (12.25)	148,892,115 (87.75)	
No job	3,147,752 (4.68)	64,129,315 (95.32)		7,486,229 (9.78)	69,032,831 (90.02)	
Frequency of computer usage			—[b]			<.001
Never	—	—		613,413 (1.31)	46,114,456 (98.71)	
Some days	—	—		1,248,875 (4.84)	24,531,551 (95.16)	
Most days	—	—		1,587,924 (11.62)	12,072,018 (88.38)	
Every day	—	—		24,802,396 (15.54)	134,817,997 (84.46)	
Frequency of internet usage			—			<.001
Never	—	—		565,220 (1.29)	43,402,368 (98.71)	

Variables	2009			2018		
	Yes	No	*P* value	Yes	No	*P* value
Per day	—	—		22,391,668 (14.56)	131,380,588 (85.44)	
Per week	—	—		4,275,620 (11.22)	33,817,721 (88.78)	
Per month	—	—		61,283 (2.47)	2,415,606 (97.53)	
Per year	—	—		260,601 (15.30)	1,442,600 (84.70)	
Enabling factors, weighted n (%)						
Family income (US $)			<.001			<.001
<35,000	1,752,244 (2.51)	68,111,146 (97.49)		3,354,653 (5.79)	54,557,513 (94.21)	
35,000-49,999	1,494,075 (4.91)	28,937,754 (95.09)		2,301,726 (9.35)	22,311,049 (90.65)	
45,000-74,999	2,476,437 (6.20)	37,486,119 (93.80)		4,179,405 (11.07)	33,558,936 (88.93)	
75,000-99,999	2,138,734 (8.22)	23,884,329 (91.78)		4,383,932 (14.32)	26,227,420 (85.68)	
≥100,000	4,749,518 (10.66)	39,788,717 (89.34)		12,334,689 (17.02)	60,133,071 (82.98)	
Health insurance coverage			<.001			<.001
Covered	12,790,369 (6.93)	171,671,557 (93.07)		27,309,470 (12.42)	192,643,180 (87.58)	
Not covered	516,217 (1.29)	39,527,626 (98.71)		927,280 (3.65)	24,489,010 (96.35)	
Needs factors , weighted n (%)						
Perceived health status			.25			<.001
Excellent	3,682,424 (5.74)	60,522,896 (94.26)		6,476,485 (9.54)	61,394,808 (90.46)	
Very good	4,582,321 (6.38)	67,271,787 (93.62)		10,517,116 (12.88)	71,134,469 (87.12)	
Good	3,539,990 (5.90)	56,413,051 (94.10)		7,737,178 (11.77)	57,995,479 (88.23)	
Fair	1,190,494 (5.46)	20,595,996 (94.54)		2,864,476 (11.81)	21,392,785 (88.19)	
Poor	324,648 (4.41)	7,033,584 (95.59)		713,007 (10.20)	6,278,595 (89.80)	
Mobility limitations			.07			.20
Yes	565,452 (4.55)	11,866,412 (95.45)		1,645,744 (10.36)	14,233,084 (89.64)	
No	12,754,425 (5.99)	200,021,362 (94.01)		26,662,518 (11.56)	204,004,900 (88.44)	
Prescribed medication by doctor or **health professional**			—			<.001
Yes	—	—		25,516,673 (15.77)	136,339,307 (84.23)	
No	—	—		2,777,904 (3.29)	81,750,890 (96.71)	
Bought drugs from another country to save money			—			<.001
Yes	—	—		745,833 (19.10)	3,159,455 (80.90)	
No	—	—		27,555,721 (11.36)	215,095,537 (88.64)	

[a]The "n" here is the number of the respondents, and the "N" here is the weighted population of the respondents.

[b]The following variables were not recorded in the 2009 survey: frequency of computer usage, frequency of internet usage, prescribed medication by doctors or health professionals, and drugs bought from another country to save money.

Results From the Multivariable Regression Model

Table 2 shows the results of a multivariable logistic regression model to evaluate the association between respondent characteristics and web-based prescription-filling behavior among respondents in 2018. After adjustment, respondents aged between 35 and 74 years, female, higher education level, and with health insurance coverage were significantly associated with web-based prescription-filling behavior. In this adjusted model, Hispanic and non-Hispanic Black individuals were found to be significantly less likely to fill the prescription on the web than non-Hispanic White individuals. Persons living in the Midwest, South, and West were more likely to fill prescriptions using web-based pharmacies than those living in the Northwest of the United States.

Table 2. Factors associated with web-based prescription-filling behavior among adults in the United States in 2018: results from the adjusted multivariable logistic regression model.

Variables	Year (2018), odds ratio (95% CI)
Predisposing factors	
Age (years)	
18-34	Ref[a]
35-49	1.27 (1.08-1.50)
50-64	1.22 (1.03-1.44)
65-74	1.38 (1.13-1.68)
75+	0.89 (0.68-1.17)
Gender	
Male	Ref
Female	1.38 (1.23-1.55)
Race	
Non-Hispanic White	Ref
Hispanic	0.71 (0.57-0.89)
Non-Hispanic Black	0.72 (0.56-0.91)
Non-Hispanic Asian	0.83 (0.63-1.09)
Non-Hispanic others	0.69 (0.36-1.30)
Education	
Less than high school	Ref
High school	1.40 (1.04-1.89)
College	1.64 (1.21-2.21)
Higher than college	2.26 (1.70-3.01)
Region	
Northwest	Ref
Midwest	1.51 (1.24-1.84)
South	1.30 (1.09-1.56)
West	1.67 (1.36-2.04)
Marital status	
Married or living with partner	Ref
Divorced, separated, widowed	0.86 (0.74-1.00)
Single, never married	0.96 (0.81-1.14)
Work status (past 12 months)	
Had job	1.00 (0.87-1.15)
No job	Ref
Frequency of computer usage	
Never	Ref
Some days	1.62 (1.06-2.50)
Most days	2.65 (1.63-4.33)
Every day	3.65 (2.38-5.62)
Frequency of internet usage	
Never	Ref
Per day	3.48 (2.11-5.73)

Variables	Year (2018), odds ratio (95% CI)
Per week	3.24 (1.97-5.34)
Per month	1.20 (0.41-3.52)
Per year	3.26 (1.45-7.35)
Enabling factors	
Family income (US $)	
<35,000	0.78 (0.65-0.93)
35,000-49,999	1.02 (0.84-1.23)
45,000-74,999	Ref
75,000-99,999	1.12 (0.94-1.35)
≥100,000	1.26 (1.08-1.48)
Health insurance coverage	
Covered	1.60 (1.19-2.14)
Not covered	Ref
Needs factors	
Perceived health status	
Excellent	Ref
Very good	1.26 (1.09-1.46)
Good	1.39 (1.19-1.63)
Fair	1.80 (1.44-2.24)
Poor	1.81 (1.27-2.58)
Mobility limitations	
Yes	1.23 (0.96-1.58)
No	Ref
Prescribed medication by doctor or **health professional**	
Yes	4.74 (3.83-5.87)
No	Ref
Bought drugs from another country to save money	
Yes	2.10 (1.51-2.93)
No	Ref

[a]Ref: reference value.

Furthermore, the frequency of computer usage was also a significant predictor of web-based prescription-filling behavior. For example, respondents who used a computer for some days, most days, and every day were all significantly more likely to report that they filled a prescription on the web. Compared with people with excellent health status, individuals with very good, good, fair, and poor health status were 1.26, 1.39, 1.80, and 1.81 times more likely to fill the prescription on the web, respectively. In addition, the probability of filling the prescription on the web was 2.10 times greater in people who had bought drugs from another country to save money. Finally, unlike the results of the unadjusted analysis, marital status and work status were not found to be significantly associated with web-based prescription-filling behavior after adjustment.

Discussion

Principal Findings

To our knowledge, this is the most recent study to investigate the trends and patterns of web-based prescription-filling behavior among a nationally representative population of adults in the United States. The results showed that the prevalence of web-based prescription filling has significantly increased from 2009 to 2018. Individuals who filled a prescription on the web were more often between the ages of 35 and 74 years, female, White, of higher education, used the computer or internet frequently, had insurance coverage, and had poorer health status.

The findings from the survey of HINTS showed that the trend of obtaining medications or vitamins on the web increased from 2003 to 2013 [7]. Our study supported the findings from the

HINTS survey and provided updated results of web-based prescription-filling behavior among adults in the United States.

In our study, characteristics including age, gender, race, education, income, and insurance coverage were found to be associated with filling prescriptions using web-based pharmacies. With regard to age, when compared with people aged between 18 and 34 years, people aged between 35 and 74 years were significantly more likely to fill a prescription on the web. The oldest group, those aged 75 years and older, was the least likely to fill a prescription on the web. There is a steady decrease in the prevalence of internet use by age. A previous study that investigated internet use among older adults showed that 82% of seniors aged 65-69 years used the internet, and the prevalence diminished with age with 60% of those aged 65-79 years and 44% of those aged ≥80 years saying they used the internet [34]. It is not surprising that older adults were least likely to fill their prescriptions on the web compared with other age groups, as these individuals usually have difficulties in adapting to technology, which makes web-based prescription filling less likely to become an alternative to obtaining prescription medications in person.

Gender is a strong predictor of web-based prescription-filling behavior. We found that the prevalence of web-based prescription-filling behavior was consistently higher in women than in men from 2009 to 2018. This may be because of the tendency of females to adapt to behaviors of web-based health technology use [35-39] and partially explained by the gender difference in health perceptions and health service utilization [35-39].

Those with higher income were also found to be more likely to fill a prescription on the web. Higher income enables people to access the computer and internet more easily, which makes them more capable of filling prescriptions on the web. Insurance coverage was also found to be associated with a higher likelihood of web-based prescription filling. A survey conducted in 1797 web-based pharmacy users and their reasons for use found that people were encouraged or even compelled to use mail-order pharmacies because of the requirement of certain health insurance plans [21]. This might explain why health insurance coverage was significantly associated with web-based prescription filling in this study.

Educational status was found to be associated with web-based prescription-filling behaviors. People with higher education were significantly more likely to fill prescriptions on the web. People with higher education often have higher health literacy, which in turn may help them access the internet more easily [40,41]. Better health literacy with higher education enables these people to understand health information easily and choose an alternative and convenient way to obtain prescriptions. In addition, higher education is linked to higher income and increases the likelihood of having a job that provides health insurance with a high-deductible insurance plan to encourage beneficiaries to fill a prescription on the web.

People with poor health status were found to be more likely to fill a prescription on the web. Poor health status could potentially result from having multiple health conditions. These individuals may have many prescriptions, which increases the complexity of the medication treatment. Web-based prescription-filling behavior provides the convenience of pharmacy access, which may explain the association between poor health status and a greater likelihood of filling a prescription using a web-based pharmacy. Previous research shows a varied relationship between health status and web-based pharmacy use. A greater burden of illness, as measured by the Charlson comorbidity index, was associated with a greater likelihood of using a web-based pharmacy [8,18], whereas another study using a health status scale similar to the one used in this study found no association [8]. Future research is needed to further understand the association between perceived health status and the behavior of web-based medication filling.

We also found that people who bought medications from another country to save money were twice as likely to fill their prescriptions on the web when compared with people who reported that they never bought medication from another country to save money. Our findings were consistent with a study from the Canadian International Pharmacy Association survey, which reported that saving money was the main reason that people purchased medications on the web, and the purchases were mostly from foreign websites [13]. Continuously rising prescription medication costs in the United States is one reason for the use of web-based pharmacies [11].

Clinical Implications

From a clinical perspective, the safety of web-based prescriptions remains to be a major concern. A significant number of unqualified and illegitimate pharmacies are active on the web [14]. Furthermore, it is difficult to distinguish between a legal and illegal web-based pharmacy, which makes the safety of purchasing medications on the web even more of an issue. In addition, a previous study showed that few web-based pharmacy users disclose their use to health care providers [8]. One important reason might be that health care providers did not ask their patients about the experience of web-based pharmacy use [42]. The findings from this study confirm the need for health care providers to be aware of the increased and prevalent use of web-based prescription-filling behavior. In addition, health care providers should actively inquire, discuss, and educate their patients on ways to identify legal and qualified web-based pharmacies [8,17,43]. The advantages and convenience of the internet will make web-based pharmacy use even more prevalent in the near future.

Web-based prescription filling may be more in demand during the current COVID-19 pandemic and future crises. Epidemic prevention policies, such as quarantines and social distancing, can make the traditional way of filling prescriptions difficult or even impossible [25]. For example, a reduction in access to traditional pharmacies may occur because of pharmacy closures or quarantines of pharmacy staff, which could drive people to seek an alternative and convenient way, such as web-based pharmacies, to fill their prescriptions. Furthermore, different patient groups may be disproportionately impacted. On the basis of our findings, older people, minority people, and people with low income and education were less likely to fill a prescription on the web. Future studies should focus on further determining the limitations of accessing web-based pharmacies and

developing appropriate interventions. The use of web-based prescription filling with mail order or home delivery among the underserved population during the pandemic crises may be one answer.

Strengths

The strengths of this study include the use of data from the NHIS, which included a large, nationally representative sample of the US noninstitutionalized population. This database allowed us to update the trend of web-based prescription filling for a decade (2009-2018) and to identify characteristics associated with web-based prescription filling in a large cohort.

Limitations

However, our study has several potential limitations. First, because the NHIS is a self-reported survey of respondents, recall bias can exist. Second, the NHIS asked respondents if they filled a prescription on the web but did not gather detailed information on the medication filled in this manner. There was no information about the class and type of medicine as well as reasons for filling a prescription on the web. Future research detailing the specific medications filled using web-based pharmacies, including prescription and nonprescription products, is warranted to further characterize patients' use of web-based pharmacies. Third, the medication distribution channels may be disrupted during the pandemic. Patients with certain medication needs, such as patients enrolled in narcotic treatment programs who need methadone may have difficulty in maintaining their treatment. Although filling methadone using web-based sources is not an option at this time, further research is warranted to evaluate the implications of the pandemic on the drug dispensing and delivery system among certain groups of patients. Fourth, due to the nature of the survey data set and study design, only the association between predictor and outcome variables can be made. Fifth, we did not investigate the implication of disease condition on the web-based prescription-filling behavior because the NHIS has only selected disease states and not all disease states. However, disease condition and severity can affect the behavior of filling a prescription on the web. An individual with a certain disease condition may be more likely to fill a prescription on the web than an individual with a different disease condition. Future research needs to investigate the association between web-based prescription-filling behavior and disease condition or severity by collecting primary data. Finally, confounding factors may exist, specifically regarding insurance coverage. For example, a patient's likelihood of filling a prescription on the web can result from his or her insurance coverage having lower copays or out-of-pocket expenses when prescriptions are filled on the web.

Conclusions

In conclusion, our findings indicate that the prevalence of web-based prescription-filling behavior among US adults has increased significantly from 2009 to 2018. Adults in the United States who were aged between 35 and 74 years, were female, were White, had higher education, were frequent users of the computer or internet, and had higher income, insurance coverage, and poorer health status were significantly more likely to fill a prescription on the web. Health care providers should be aware of the increased use of web-based pharmacies to fill prescriptions and help educate patients about the appropriate use of legitimate web-based pharmacies.

Acknowledgments

This study is based in part on data from the NHIS database provided by the NCHS and managed by the CDC. The interpretation and conclusions contained herein do not represent those of the NCHS or the CDC. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The results were presented in part at the 2019 International Society for Pharmacoepidemiology (ISPE) ISPE's 12th Asian Conference on Pharmacoepidemiology (Kyoto, Japan, October 2019).

Conflicts of Interest

None declared.

References

1. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
2. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
3. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. BMC Public Health 2007;7:53 [FREE Full text] [doi: 10.1186/1471-2458-7-53] [Medline: 17425798]
4. Fox SD. Health Online 2013: One in three American adults have gone online to figure out a medical condition. Pew internet Project and California Health Care Foundation: Pew internet Project and California Health Care Foundation; 2013. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2021-02-10]
5. Wong DK, Cheung M. Online Health Information Seeking and eHealth Literacy Among Patients Attending a Primary Care Clinic in Hong Kong: A Cross-Sectional Survey. J Med Internet Res 2019 Mar 27;21(3):e10831 [FREE Full text] [doi: 10.2196/10831] [Medline: 30916666]

6. Rupp MT. Attitudes of Medicare-eligible Americans toward mail service pharmacy. J Manag Care Pharm 2013 Sep;19(7):564-572. [doi: 10.18553/jmcp.2013.19.7.564] [Medline: 23964618]
7. Health Information National Trends Survey (HINTS). National Cancer Institute. 2003. URL: https://hints.cancer.gov/view-questions-topics/question-details.aspx?red=1&qid=759 [accessed 2021-02-10]
8. Desai K, Chewning B, Mott D. Health care use amongst online buyers of medications and vitamins. Res Social Adm Pharm 2015 Nov;11(6):844-858. [doi: 10.1016/j.sapharm.2015.01.001] [Medline: 25769500]
9. U.S. Food and Drug Administration. 2018. URL: https://www.fda.gov/drugs/besaferx-know-your-online-pharmacy/media#7 [accessed 2021-02-10]
10. Henney JE. Cyberpharmacies and the role of the US Food And Drug Administration. J Med Internet Res 2001;3(1):E3. [doi: 10.2196/jmir.3.1.e3] [Medline: 11720945]
11. Observations on Trends in Prescription Drug Spending.: Department of Health and Human Services; 2016. URL: https://aspe.hhs.gov/system/files/pdf/187586/Drugspending.pdf [accessed 2021-02-10]
12. Roger BG. Aparna Mathur. In Whom We Trust: The Role of Certification Agencies in Online Drug Markets. The BE Journal of Economic Analysis & Policy 2013;14(1):50. [doi: 10.1515/bejeap-2013-0085]
13. Smith T. CIPA survey of Americans finds 64% purchase medications from Canada due to affordability. Canadian International Pharmacy Association: Canadian International Pharmacy Association; 2015. URL: https://www.cipa.com/cipa-survey-of-americans-finds-64-purchase-medications-from-canada-due-to-affordability/ [accessed 2021-02-10]
14. US Food and Drug Administration. 2018. URL: https://www.fda.gov/consumers/consumer-updates/how-buy-medicines-safely-online-pharmacy [accessed 2021-02-10]
15. The National Association of Boards of Pharmacy. Internet Drug Outlet Identification Program Progress Report for State and Federal Regulatorsptember 2018. URL: https://nabp.pharmacy/wp-content/uploads/2018/09/Internet-Drug-Outlet-Report-September-2018.pdf [accessed 2021-02-15]
16. Szekely PK, Fittler A, Hancu G, Hajnal F, Ciurba A, Sipso E, et al. Attitude Of patients and customers regarding purchasing drugs online. Farmacia 2015;63(1):93-98.
17. Fittler A, Lankó E, Brachmann B, Botz L. Behaviour analysis of patients who purchase medicines on the internet: can hospital pharmacists facilitate online medication safety? Eur J Hosp Pharm 2012 Sep 07;20(1):8-12. [doi: 10.1136/ejhpharm-2012-000085]
18. Brown J, Li C. Characteristics of online pharmacy users in a nationally representative sample. J Am Pharm Assoc (2003) 2014 May;54(3):289-294. [doi: 10.1331/JAPhA.2014.13169] [Medline: 24816356]
19. Fittler A, Vida RG, Káplár M, Botz L. Consumers Turning to the Internet Pharmacy Market: Cross-Sectional Study on the Frequency and Attitudes of Hungarian Patients Purchasing Medications Online. J Med Internet Res 2018 Aug 22;20(8):e11115 [FREE Full text] [doi: 10.2196/11115] [Medline: 30135053]
20. Abanmy N. The extent of use of online pharmacies in Saudi Arabia. Saudi Pharm J 2017 Sep;25(6):891-899 [FREE Full text] [doi: 10.1016/j.jsps.2017.02.001] [Medline: 28951675]
21. Mazer M, DeRoos F, Shofer F, Hollander J, McCusker C, Peacock N, et al. Medications from the web: use of online pharmacies by emergency department patients. J Emerg Med 2012 Feb;42(2):227-232. [doi: 10.1016/j.jemermed.2010.05.035] [Medline: 20656438]
22. Alfahad N, Albelali M, Khurshid F, Al-Arifi M, Al-Dhawailie A, Alsultan M. Perception and Knowledge to Online Pharmacy Services among Consumers in Riyadh, Saudi Arabia: a Pilot Survey. Latin American Journal of Pharmacy 2015:a [FREE Full text] [doi: 10.4103/0973-8258.82090]
23. LegitScript: Internet Pharmacy Policy Guide, 2015. Country by Country Internet Pharmacy Regulations. URL: https://tinyurl.com/b24udpjk [accessed 2021-02-16]
24. Clement J. Global number of internet users. Statista: Statista URL: https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/ [accessed 2021-02-10]
25. Alexander GC, Qato DM. Ensuring Access to Medications in the US During the COVID-19 Pandemic. JAMA 2020 Jul 07;324(1):31. [doi: 10.1001/jama.2020.6016]
26. National Center for Health Statistics. URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2018/srvydesc.pdf [accessed 2021-02-21]
27. National Center for Health Statistics. 2017. URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2016/srvydesc.pdf [accessed 2021-02-21]
28. National Center for Health Statistics. NHIS (PUDRD). 2015. URL: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2015/srvydesc.pdf [accessed 2021-02-21]
29. Cohen RA, Adams PF. Use of the internet for health information: United States, 2009. NCHS Data Brief 2011 Jul(66):1-8 [FREE Full text] [Medline: 22142942]
30. Andersen RM. National health surveys and the behavioral model of health services use. Med Care 2008 Jul;46(7):647-653. [doi: 10.1097/MLR.0b013e31817a835d] [Medline: 18580382]
31. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? J Health Soc Behav 1995 Mar;36(1):1-10. [Medline: 7738325]

32. Andersen R, Newman JF. Societal and individual determinants of medical care utilization in the United States. Milbank Mem Fund Q Health Soc 1973;51(1):95-124. [Medline: 4198894]
33. Varma NK. Simplifying the Analysis of Complex Survey Data Using the SAS® Survey Analysis Procedures. Orlando, Florida. 2012. URL: https://support.sas.com/resources/papers/proceedings12/348-2012.pdf [accessed 2021-02-10]
34. Anderson MP. Tech Adoption Climbs Among Older Adults. Pew Research Center: Pew Research Center; 2017. URL: https://www.pewresearch.org/internet/2017/05/17/tech-adoption-climbs-among-older-adults/ [accessed 2021-02-10]
35. Bertakis KD, Azari R, Helms LJ, Callahan EJ, Robbins JA. Gender differences in the utilization of health care services. J Fam Pract 2000 Feb;49(2):147-152. [Medline: 10718692]
36. Hibbard JH, Pope CR. Gender roles, illness orientation and use of medical services. Soc Sci Med 1983;17(3):129-137. [doi: 10.1016/0277-9536(83)90246-0] [Medline: 6836347]
37. Waldron I. Sex differences in illness incidence, prognosis and mortality: issues and evidence. Soc Sci Med 1983;17(16):1107-1123. [doi: 10.1016/0277-9536(83)90004-7] [Medline: 6623118]
38. Cleary PD, Mechanic D, Greenley JR. Sex differences in medical care utilization: an empirical investigation. J Health Soc Behav 1982 Jun;23(2):106-119. [Medline: 7108177]
39. Verbrugge LM, Wingard DL. Sex differentials in health and mortality. Women Health 1987;12(2):103-145. [doi: 10.1300/J013v12n02_07] [Medline: 3424846]
40. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health Literacy and Health Information Technology Adoption: The Potential for a New Digital Divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
41. Protheroe J, Whittle R, Bartlam B, Estacio EV, Clark L, Kurth J. Health literacy, associated lifestyle and demographic factors in adult population of an English city: a cross-sectional survey. Health Expect 2017 Feb;20(1):112-119 [FREE Full text] [doi: 10.1111/hex.12440] [Medline: 26774107]
42. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the Internet for medical information. J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [Medline: 11929503]
43. Fung CH, Woo HE, Asch SM. Controversies and legal issues of prescribing and dispensing medications using the Internet. Mayo Clin Proc 2004 Feb;79(2):188-194. [doi: 10.4065/79.2.188] [Medline: 14959914]

Abbreviations

CDC: Centers for Disease Control and Prevention
HINTS: Health Information National Trends Survey
ISPE: International Society for Pharmacoepidemiology
MEPS: Medical Expenditure Panel Survey
NABP: National Association of Boards of Pharmacy
NCHS: National Center for Health Statistics
NHIS: National Health Interview Survey
PSU: primary sampling unit

Edited by R Kukafka; submitted 19.08.20; peer-reviewed by A Mavragani, R De Boni; comments to author 09.10.20; revised version received 01.12.20; accepted 04.02.21; published 16.03.21.

Please cite as:
Yang LY, Lyons JG, Erickson SR, Wu CH
Trends and Characteristics of the US Adult Population's Behavioral Patterns in Web-Based Prescription Filling: National Survey Study
J Med Internet Res 2021;23(3):e23662
URL: https://www.jmir.org/2021/3/e23662
doi:10.2196/23662
PMID:33724193

Original Paper

Opportunities and Challenges for Digital Social Prescribing in Mental Health: Questionnaire Study

Shivani Patel[1], MRCPsych; Gerry Craigen[2,3], MD; Mariana Pinto da Costa[1,4,5], MD; Becky Inkster[6,7], DPhil

[1]South London and Maudsley NHS Trust, London, United Kingdom

[2]Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[3]Centre for Mental Health, University Health Network – Toronto General, Toronto, ON, Canada

[4]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[5]Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal

[6]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

[7]Finance and Economics Programme, The Alan Turing Institute, London, United Kingdom

Corresponding Author:
Shivani Patel, MRCPsych
South London and Maudsley NHS Trust
Denmark Hill
London
United Kingdom
Phone: 44 3228 6000 ext 1234
Email: shivaninpatel183@gmail.com

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/3/e29041

Abstract

Background: The concept of digital social prescription usually refers to social prescriptions that are facilitated by using technology. Tools that enable such digital social prescriptions may be beneficial in recommending nonmedical activities to people with mental illness. As these tools are still somewhat novel and emerging, little is known about their potential advantages and disadvantages.

Objective: The objective of this study is to identify the potential opportunities and challenges that may arise from digital social prescriptions.

Methods: We developed a qualitative questionnaire that was disseminated through social media (Facebook and Twitter). A purposive sample targeting *digital mental health experts* and nonexperts was approached. The questionnaire asked participants' views about digital social prescription; the core elements linked with a definition of digital social prescription; and the strengths, weaknesses, opportunities, and threats associated with digital social prescription.

Results: Four core elements were recommended to define the concept of digital social prescription: digital, facilitate, user, and social. The main strength identified was the possibility to rapidly start using digital social prescription tools, which were perceived as cost-effective. The main weaknesses were their poor adherence and difficulties with using such tools. The main opportunities were an increased access to social prescription services and the prevention of serious mental illness. The main threats were certain groups being disadvantaged, patients being subject to unintended negative consequences, and issues relating to confidentiality and data protection.

Conclusions: Although digital social prescriptions may be able to effectively augment the social prescriptions, a careful consideration of practical challenges and data ethics is imperative in the design and implementation of such technologies.

(J Med Internet Res 2021;23(3):e17438) doi:10.2196/17438

KEYWORDS

mental health; technology; psychiatry; mobile phone

Introduction

Background

The idea of health care professionals prescribing activities to their patients has been around since the 1990s when contemporary exercise referral schemes were first created [1]. The term *social prescription* has since been defined as "a means of enabling general practitioners (GPs) and other frontline health care professionals to refer people to 'services' in the community instead of offering only medical solutions" [2]. Social prescribing services are typically offered by voluntary and community sector organizations and usually involve a person who supports people to access local activities [3]. Examples of activities may range from traditional formalized programs such as smoking cessation programs to exercise, cooking classes, and befriending services [4-6]. The benefits of social prescribing have previously been explored, with studies suggesting a reduction in GP consultations and accident and emergency department attendance when social prescribing services are working well [7], and a reduced requirement for psychiatrists and mental health nurse consultations [8]. The term *digital social prescription* has previously been described as "any digital solution, technology, information or electronic system that enables social prescribing" [9].

Digital technologies have become increasingly pervasive within the society [10], and our dependence on interactive technologies for the delivery of health care has been particularly important during the global COVID-19 pandemic [11]. Interactive technologies have successfully enabled changes in human attitudes and behaviors [12,13], and the use of this technology for social prescription could offer a health benefit to our modern society. Currently, digital social prescription tools (DSPTs) used in the United Kingdom are used for patients with physical health comorbidities. DSPTs, such as those developed by Evergreen Life [14] and Elemental [15], use electronic patient records and community directory software to match nonmedical activities that have been shown to benefit a patient's medical condition. The matching process involves using an algorithm designed to match activities to a patient based on their preferences, comorbidities, and locality. This process aims to tailor nonmedical interventions to the needs and preferences of the patient in a sophisticated and efficient manner.

Objective

The objective of this study is to collect the views of both experts and the general public on digital social prescription while focusing on the core elements that should base the concept of digital social prescription and identify the potential benefits and challenges that may arise from digital social prescriptions.

Methods

Study Design

This study includes a qualitative questionnaire (Multimedia Appendix 1) with views of both experts and nonexperts on the potential use of digital social prescriptions.

Instrument

The questionnaire started with a short introduction of digital social prescriptions, including a diagram on how it might work in practice. The questionnaire asked participants' views of digital social prescription; the core elements linked with a definition of digital social prescription; and the strengths, weaknesses, opportunities, and threats (SWOT) associated with digital social prescription.

Data Sampling and Collection

The instrument targeted digital mental health experts and nonexperts. Experts were selected from a purposive sample of researchers who had published in the *Journal of Medical Internet Research* on a topic relating to digital mental health in the last 5 years and were contacted by email. Nonexperts were approached through social media platforms (Facebook and Twitter).

Data Analysis

We used content analysis [16] and the SWOT framework to analyze responses from participants. SWOT frameworks are commonly used in strategic analysis to analyze the internal (strengths and weaknesses) and external (opportunities and threats) factors relating to a project concept or idea [17]. The first author (SP) coded all the material, and the third author (MP) reviewed all the data to ensure the consistency and credibility of the coding and grouping [18].

Results

Sociodemographic Data of Participants

Our sample consisted of 22 *nonexpert* participants and 22 *expert* participants (Table 1).

Table 1. Demographics of the sample.

Demographics	Expert, n (%)	Nonexpert, n (%)
Age (years)		
0-20	0 (0)	2 (9)
20-30	12 (55)	14 (64)
30-40	8 (36)	4 (18)
40-50	2 (9)	1 (4.5)
50+	0 (0)	1 (4.5)
Gender		
Male	8 (36)	10 (45.5)
Female	69 (59)	12 (54.5)
Nonbinary	23 (5)	0 (0)
Nationality		
British	6 (27)	19 (86.5)
Canadian	4 (18)	1 (4.5)
Indian	0 (0)	1 (4.5)
Greek	0 (0)	1 (4.5)
Dutch	2 (9)	0 (0)
American	8 (41.5)	0 (0)
Australian	1 (4.5)	0 (0)
Occupation		
Mental health professional (doctor, psychologist, or mental health worker)	0 (0)	8 (36)
Student	1 (4.5)	12 (54.5)
Researcher	21 (95.5)	1 (4.5)
Unemployed	0 (0)	1 (4.5)

Definition of Digital Social Prescription

Expert and nonexpert participants were asked to provide a definition of digital social prescriptions. For both groups, the responses gathered identified four core elements: (1) digital, (2) facilitate, (3) user, and (4) social (Table 2).

As a result, the following definition is proposed: *digital social prescription refers to social prescriptions that have been facilitated through the use of technology, such as mobile phone apps or online platforms intended to benefit its users.*

We used the terms *digital social prescription tools* and *digital platforms* interchangeably to reflect the views of our participants.

The findings from our SWOT analysis are reported in Textbox 1 and Textbox 2.

Table 2. Expert participants' (N=22) and nonexpert participants' (N=22) responses to the question "How would you define digital social prescription?" grouped by core element.

Participant type	Words from participants' responses			
	Digital	Facilitate	User	Social
Expert	• Digital • Technology • Web-based platforms	• Use • Facilitate	• Recommended as a part of health care • Prescribed by a clinician • Recommended to patients	• Social prescription
Nonexpert	• Technology • App • Digital platform	• Facilitate • Link • Allow	• Self • Patients • Doctors	• Nonmedical activities • Social prescription

Textbox 1. Strengths, weaknesses, opportunities, and threats analysis of responses from expert participants (n=22).

Strengths

- Time (n=13)
  - Quick to start
  - Quick to download
- Easy to use (n=11)
  - Intuitive for users
  - Easy to use
- Social connection (n=7)
  - Social connection in local area
  - Social connection in area

Weaknesses

- Loss of interest (n=15)
  - Fatigue
  - High drop-off rate
  - Lack of continuity
- Hard to use (n=8)
  - Technical difficulties to use
  - Not acceptable to disadvantaged groups—lower socioeconomic groups, older people, and people with physical health comorbidities
- Authenticity of participation (n=1)
  - Interference from bots and trolls
- Difficulty in remaining updated (n=4)
  - Difficult to keep up with new technologies
  - Difficulties with maintaining lists

Opportunities

- Improved access (n=15)
  - Access for more people
  - Greater access if done equitably
  - Access to care for poorer groups in the society
  - Access for hard-to-reach groups—poor mobility and poor socioeconomic groups
- Loneliness (n=7)
  - Help to combat isolation
  - Target loneliness
- Resource efficiency (n=2)
  - Help to free up resources that can be redirected toward significant mental illness

Threats

- Privacy and confidentiality relating to data (n=14)
  - Privacy of data
  - Use and storage of data

- Widening the health gap (n=5)
  - Widen the gap between those who can afford technologies and those who cannot
  - Digital divide exacerbating health inequalities
- Not accepted by establishment (n=1)
  - Seen as a fad by traditional clinicians

Textbox 2. Strengths, weaknesses, opportunities, and threats analysis of responses from nonexpert participants (n=22).

Strengths

- Cost (n=17)
  - Cheap or cost-effective
  - Requires fewer human resources to be involved in the process as it uses an algorithm for matching
- Time (n=10)
  - Quick
  - Reduce the waiting time between patient expressing an interest and being able to start an activity
- User experience (n=9)
  - System more transparent for patients as they can track their social prescribing referral throughout the process
  - "On-demand" service
  - Younger generations might find it easier to engage
  - All that there is to offer in one place
- Local (n=8)
  - Able to easily identify activities close to patient's location
  - Digital social prescribing will match patient with local activities, allowing patients to feel more connected to their community
  - Easy to find available activities (n=5)
  - Update activities quickly
  - Greater variety of activities and easier to keep a register
- Efficiency (n=3)
  - Less paperwork

Weaknesses

- Difficulty in using the tool (n=19)
  - Difficulties in using it
  - Older generation and very ill patients might find it difficult to use such tools
  - Language barriers
- Poor engagement (n=14)
  - People not turning up to activities
  - People not using it over longer periods
  - Not everyone understands the intended goal
  - May not be culturally appropriate
- Lack of human connection (n=13)
  - Patients feel they are not being listened to
  - Patients might be distrustful, lack of link worker to help with building trust
  - No function for support workers to provide guidance
  - Patient expectations for the management of problem (n=10)
  - "Tech solution" might put people off
  - People may feel this is not an appropriate response
  - Mismatch between patients' expectations of an activity and the reality
  - People might be offended that they are asked to use a digital app instead of being able to talk to a health care professional in the first instance

- Delay in appropriate management (n=1)
  - Delay in treatment
  - Problems with maintaining lists of activities (n=3)
  - Problems with social prescription—directories with activities are not free of errors or comprehensive
  - Community centers are paper based
  - Activities do not get listed
  - Cost of keeping this system updated

Opportunities

- Greater access (n=20)
  - Greater access to activities
  - Allow for a more widespread uptake of social prescription
- Address loneliness and social connection (n=10)
  - Improve social connections for those who are isolated
- Role in prevention (n=11)
  - Potential role in prevention of mental health disorders through strengthening social connections
  - Cheaper cost might mean rolled out earlier to help in prevention

Threats

- Patient protection from adverse unintended consequences (n=5)
  - Those providing activities may not have the patients' best interests
  - No clear way that patients are being protected from outsiders
- Confidentiality and data protection (n=10)
  - Data may be sold for profit
  - Data may not be kept safe
  - Hackers may access data
- Bias (n=10)
  - Educated middle-class groups more likely to use technology to their advantage than those who need services
  - Some groups may be favored over other either through the algorithm being inherently biased or access only being available in neighborhoods who can afford to invest in a digital solution
- Not helpful for some groups (n=5)
  - Not helpful for all mental health conditions
  - Many people are not online and do not wish to be, some of the groups who need social prescribing the most are among these

Strengths

The expert group identified the main strength of DSPTs as being quick to start, whereas the nonexpert group perceived the main strength as their potential cost-effectiveness.

Both experts and nonexperts suggested that DSPTs would be faster to use; the nonexpert group suggested that using a digital platform would make the process of social prescription faster partly through a reduction in paperwork for those prescribing the activity. Both the expert and nonexpert groups commented on DSPTs being easy to use and having an improved user experience. Nonexpert participants suggested that reasons for these included users being able to clearly track their referral through the platforms, the platforms providing an *on-demand* service, and that all activity information would be consolidated in one place. They further suggested that younger people would find this way of accessing services easier to navigate than traditional methods. Both expert and nonexpert groups also suggested that DSPTs could be used to help individuals feel more connected to their local community.

The nonexpert group suggested that cost-effectiveness would be a significant advantage of DSPTs, whereas none of the expert

participants commented on their cost-effectiveness. The nonexpert group suggested that although digital social prescription would use an algorithm for matching patients, there would be fewer people who would need to be involved in the social prescription process, which may result in the process being less costly.

Weaknesses

Experts identified the main weakness of DSPTs as having a high dropout rate, whereas nonexpert groups were concerned that certain groups would find technology particularly difficult to use.

Both experts and nonexperts commented on the loss of interest and high dropout rates of patients using DSPTs. One expert suggested that patients may be fatigued with technology *solving* problems, and nonexperts additionally suggested that patients may not understand the point of DSPTs and may therefore not be motivated to continue using it. Both experts and nonexperts identified that DSPTs may be difficult for certain groups to use. These groups included older people, people with physical health disabilities, people from lower socioeconomic groups, and people with cultural or language barriers. Both experts and nonexperts also commented on the difficulty of maintaining the updated lists of local activities.

Experts commented on specific issues related to the technology used in facilitating digital social prescriptions. Experts commented on the difficulty in health care services being able to keep up with new developments in technology. They also commented on the potential interference on platforms by bots and trolls, which may affect the authenticity of participation.

Nonexperts raised concerns about DSPTs being inappropriate for those experiencing serious mental illness or where activities on offer may not be culturally appropriate. Several participants commented on digital social prescriptions resulting in a possible loss of human connection, perceived as inappropriate by patients and their families. A delay in appropriate treatment has also been cited as a potential weakness.

Opportunities

Both experts and nonexperts felt that the main opportunity relating to digital social prescription was an increased access to mental health care. Experts particularly felt that this may be of particular benefit to *hard-to-reach* groups, including those from poorer socioeconomic backgrounds or those with other physical health comorbidities. Both experts and nonexperts perceived DSPTs as a potential help to prevent loneliness and improve social connection.

One expert commented on digital social prescription helping with resource efficiency by freeing up resources that could be directed to those experiencing significant mental illness. Nonexperts considered DSPTs to play a role in the prevention of mental health disorders.

Threats

Both experts and nonexperts were concerned with data protection, confidentiality, and the potential monetization of data. Both experts and nonexperts also commented on the potential of bias resulting in a widening of health outcomes among different groups of individuals. This may be due to affluent middle-class individuals being the early adopters of new technology or due to the algorithms used in the DSPTs being inherently biased against certain groups. Nonexpert participants also commented that digital social prescriptions may be funded in certain areas, but this may not be the case in other areas.

Nonexperts considered that some individuals who would benefit from social prescription may not want to use new digital technologies to access activities. They also note that digital social prescriptions may not be beneficial for all mental health conditions. Some participants expressed concern regarding unintended consequences of digital social prescription; for example, if the activity involved patients volunteering at a coffee shop, then these patients may be exploited as free labor.

Experts additionally suggested that digital social prescriptions may be seen as a fad by clinicians and rejected.

Discussion

Principal Findings

From the consultation of the various participants, our study proposes a definition for digital social prescription: “Digital social prescription refers to social prescription that has been facilitated through the use of technology, such as mobile phone apps and online platforms intended to benefit users.”

The main perceived benefits of DSPTs were improved access to mental health care, fast adoption by users, and cost-effectiveness. Other perceived benefits included improved user experience, helping users feel more connected to their local communities, and potential prevention of loneliness and serious mental illness. There appeared to be significant crossover with regard to the perceived benefits of DSPTs from both experts and nonexperts. The main exception to this was cost-effectiveness, which was considered a significant benefit from nonexperts but was not commented on by the expert group.

The main challenge of DSPTs identified from our questionnaire was a poor engagement with such tools and certain groups finding the technology difficult to use. Other challenges include the DSPT being viewed as inappropriate by both patients and clinicians, certain groups being effectively excluded from using these tools, unintended negative consequences for patients, and concerns with confidentiality and data protection. Experts also commented on the difficulty faced by health care providers in keeping up with developments in technology and security, which may include issues relating to data hacking and interference from artificial intelligence–powered bots or trolls. The responses to the potential challenges from DSPT between experts and nonexperts were broadly similar; however, experts emphasized the challenges of technology more than nonexperts. Interestingly, almost all the expert respondents also commented on the high dropout rates of DSPTs, which may reflect their own experiences from working in the field and their concerns with user engagement.

Strengths and Limitations

To our knowledge, this is the first qualitative study to explore the potential benefits and challenges of digital social prescription and suggests a definition of digital social prescribing that originated from such views. Our study compared the responses of a purposively selected sample of experts in digital mental health with those of nonexperts. The overall sample included a range of different ages, genders, nationalities, and occupations. By comparing the views of experts with nonexperts, we were able to identify key similarities and differences in their perspectives and views on digital social prescription, which, for the most part, were largely similar.

The main limitation of this study was that it had a small sample size. In addition, while focusing on incorporating views of experts and nonexperts (ie, the general public), there might have been other stakeholders, such as clinicians, patients, and caregivers, who we have not particularly targeted in this study. This would be an important area of further research, particularly as the use of digital tools in health care has become more prevalent. It is also worth noting that none of the expert cohort were older than 50 years, which may skew the views provided.

Comparison With the Literature

The discussion of these findings was organized to reflect the themes that emerged in our study. The themes that were mentioned most frequently are discussed first.

It is important to note that as the majority of studies relating to social prescription refer to nonpharmacological prescription of exercise (exercise groups, gym programs, etc), most of the available literature concerns nonmental health–specific social prescription programs. Nevertheless, they provide a basis for understanding some of the core discussions regarding implementation and barriers to social prescription, which may also be relevant to digital social prescribing for mental health.

A key benefit of DSPTs identified by both experts and nonexperts was improved access to mental health care. Access to mental health care is a significant issue worldwide. The Five Year Forward View of Mental Health published in 2016 identified that approximately 15% of those with anxiety and depression were being seen by Improving Access to Psychological Therapy services [19]. The provision of services in low- and middle-income countries is even more sparse, with estimates suggesting that up to 90% of individuals living with mental health disorders are receiving no mental health care [20]. Access to smartphones has been a global phenomenon, and there has been a considerable interest in delivering mental health care through mobile phone technology [21]. Young people have been shown to adopt new technologies quickly and to use mobile phone technology in the event of sickness, personal health crises, or in response to health concerns of others [22]. In the United Kingdom, a majority of mental health conditions are managed through primary care, and it has been suggested that the use of technology may allow for more options of self-referral with automated or semiautomated interventions, thereby improving access [23].

Cost-effectiveness was perceived as one of the main benefits of digital social prescriptions by nonexpert participants in this study. Cost-effectiveness and social prescriptions have been a hotly debated topic over the past decade. Some studies have indicated that social prescription may result in fewer hospital and GP appointments, thereby translating into reduced costs for the National Health Service [7]. However, critics have suggested that there is a poor evidence for sustained improved health care outcomes [24,25] and that social prescription programs that have demonstrated positive health outcomes incur a higher cost than traditional care [26]. A systematic review of physical activity interventions in primary care showed that interventions ranged from £304 (US $425) to £75,982 (US $106,346) per quality-adjusted life year depending on the scheme intensity [27]. Digital social prescription may provide a greater efficiency in some respects to matching individuals with activities, but if the bulk of the cost depends on how individual programs are run, then the use of a digital platform may only have a marginal effect on costs for social prescription programs.

One of the main barriers in assessing social prescription programs is that the programs delivered by third-sector organizations often have limited funding, and it is therefore difficult to gather data on outcomes over a sustained period [28]. It is likely that this same problem will exist with digital social prescribing programs, as the activities that are matched with patients would also be largely provided through third-sector organizations.

Interestingly, the study participants did not comment on the intrinsic benefits and functionality that technology may have beyond being quick and easy to use. A review conducted by Husk et al [29] did not identify speed and efficiency as important factors leading to the successful use of social prescription programs, and human factors such as support from their link worker or practical support, such as free travel for activities, mattered much more to participants. There may, however, be opportunities provided by using digital means to access social prescriptions. Hollis et al [30] described the potential of mobile phone apps having embedded validated measures such as the Patient Health Questionnaire-9 depression scale as well as the option for patients to track their symptoms over time. With respect to DSPTs, this may also mean that large amounts of user data that can be used to evaluate the effectiveness of these tools can be collected quickly and accurately.

Adherence to DSPTs was identified as the main challenge by both experts and nonexperts. Indeed, adherence to traditional social prescription programs has also been shown to be challenging. Pavey et al [31] conducted a systematic review of the uptake and adherence to exercise referral schemes, which are the most common social prescription in the United Kingdom. They identified that the pooled level of adherence to exercise referral schemes was only 49% in observational studies and 43% in randomized controlled trials. In studies examining factors that improve adherence to social prescription programs, the relationship between navigators and patients has been shown to be one of the most important factors facilitating social prescription [32,33]. The skill of those conducting the activity also appears to be an important factor for adherence [34,35] as well as patients being able to see positive results from undertaking activities [36]. Given the existing literature, one can assume that a purely digital social prescription platform, in

which there is no direct human contact, may result in even poorer adherence. However, a digital platform may allow participants to record key data, such as sleep and mood, and improvements in these parameters may improve adherence.

Several barriers to using digital social prescriptions were also described. Cultural and religious factors are likely to play an important role in determining whether a social prescription will be effective. In several cultures, seeking help for mental health conditions can often be stigmatizing [37], and some activities such as mixed-sex swimming may be seen as inappropriate in the context of an individual's culture. Language may also be a significant barrier in allowing individuals to participate in a prescribed activity if the DSPT is only available in English. In addition, digital barriers were also described by participants. Older people in the United Kingdom have been shown to experience high rates of loneliness as compared with other groups in the society [38]; however, official Office for National Statistics data in 2019 showed that from those aged ≥75 years who participated in the survey, less than half used the internet [39]. Ethnicity has also been shown to contribute to the digital divide, with studies showing that Black, minority, and ethnic backgrounds are more likely to access computers outside their own homes as compared with White individuals [40]. This brings with it the challenge of ensuring adequate privacy in engaging with internet-based content related to an individual's mental health.

Data protection and information sharing are important factors to consider in digital social prescriptions. It also appears to be a concern for consumers. In a 2017 survey, confidence in the data security of technology companies declined from 31% in 2016 to 24% in 2017 [41]. Confidentiality is an important tenet of medicine; however, in practice, there are many scenarios in which information sharing between parties is necessary to provide the best care for a patient. Guidelines relating to social prescription have indicated that it is the responsibility of the referrer to transfer any relevant information to the person conducting the nonmedical activity [42]. Despite this, survey data [41] have indicated that patients are much more averse to sharing their data with nonphysicians, even if these parties are integral to the delivery of patient care. Clear guidelines explaining how data are used and stored would be required to ensure that the consent from patients is valid. It would also be necessary to consider how these security rules would be enforced and what remedies should be offered to those affected by security breaches.

Algorithmic programming is central to the apps that we use today and is likely to be used in the development of a DSPT. These algorithms might result in potential race discrimination, gender discrimination, and ageism [43,44]. This may also be an important consideration with regard to a DSPT. Existing psychiatric risk assessment tools that have been shown to have poor accuracy [45] may be integrated into digital social prescribing software, further resulting in an effective discrimination against certain groups. Furthermore, clinicians who may be involved in designing these tools may introduce their own biases, which could include greater patient restrictions, particularly for those of certain ethnic backgrounds [46]. Organizations, including the Open Data Institute, are considering the potential ethical implications arising from the use of digital tools and have suggested the use of ethical frameworks such as the Data Ethics Canvas [47] to address these issues.

Although there have been no known studies directly looking at the unintended consequences of digital social prescription, bridging the online and offline worlds can create risks, and in cases where things might go wrong, liability may be an issue for both clinicians and software developers. There has been some discussion of the potential negative consequences relating to social prescription [48], which includes patients becoming stressed by the commitment required or becoming so consumed in an activity that they neglect other key aspects of their life and well-being.

Implications for Practice, Research, and Policies

DSPTs may be a helpful method for delivering nonmedical activities to those with mental illness. There are various types of DSPTs with their differences, although with a commonality of providing patients with nonmedical activities that are available in a patient's local area. The use of such DSPTs may result in greater accessibility of activities for patients and may be more cost-effective than traditional social prescription methods.

There are several challenges associated with digital social prescriptions. First, digital social prescriptions may not be appropriate for all patients. A careful consideration of symptomatology and patient expectations must be considered before making any universal recommendations. Barriers to using digital social prescriptions are likely to exist. This may include cultural and language barriers, difficulty with using the technology due to unfamiliarity, or difficulty with using the platform due to physical impairment. Cost may also be a prohibitive factor. These barriers need to be studied in more detail, and steps should be taken to improve access to digital social prescriptions. Issues relating to patient confidentiality and data protection are likely to arise in the development of DSPTs. These issues should be considered at every stage of the development and implementation of digital social prescription programs.

Although digital social prescriptions may be of benefit to patients, there is not enough evidence to substantiate this. Research looking at short-term and long-term outcome measures, such as clinical impact and cost-effectiveness, is required to identify the true benefit. Given that adherence to DSPTs was identified as the main perceived challenge, research into how adherence may be improved would also be important. On the basis of the data collected from this research, decisions can be made as to whether DSPTs should be used more widely in mental health care.

Conclusions

Digital social prescriptions may be able to provide important opportunities and help to reduce the burden of distress in patients. Important patient considerations ranging from the appropriateness of an activity to patient discrimination will need to be carefully considered in the design and implementation of this technology. More evidence is needed to further support the advancement of digital social prescribing, but with more

rigorous research and respect for data ethics, this may be a significant advancement in 21st century medicine.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Digital social prescribing questionnaire used in the study.
[DOCX File , 332 KB - jmir_v23i3e17438_app1.docx]

References

1. Pavey TG, Taylor AH, Fox KR, Hillsdon M, Anokye N, Campbell JL, et al. Effect of exercise referral schemes in primary care on physical activity and improving health outcomes: systematic review and meta-analysis. BMJ 2011 Nov 06;343(nov04 2):d6462-d6462 [FREE Full text] [doi: 10.1136/bmj.d6462]
2. Social Prescribing. NHS England. 2018. URL: https://www.england.nhs.uk/personalisedcare/social-prescribing/ [accessed 2018-11-04]
3. What is social prescribing? 2018. The Kings Fund. URL: https://www.kingsfund.org.uk/publications/social-prescribing/ [accessed 2018-11-04]
4. Kimberlee R. What is social prescribing? ASSRJ 2015 Jan 25;2(1) [FREE Full text] [doi: 10.14738/assrj.21.808]
5. Jones M, Kimberlee R, Deave T, Evans S. The Role of Community Centre-based Arts, Leisure and Social Activities in Promoting Adult Well-being and Healthy Lifestyles. IJERPH 2013 May 10;10(5):1948-1962. [doi: 10.3390/ijerph10051948]
6. Pinto da Costa M. Can social isolation caused by physical distance in people with psychosis be overcome through a Phone Pal? Eur. Psychiatr 2020 May 22;63(1):1-3. [doi: 10.1192/j.eurpsy.2020.53]
7. Polley M, Pilkington K. A review of the evidence assessing impact of social prescribing on healthcare demand and cost implications. University of Westminster. URL: https://westminsterresearch.westminster.ac.uk/item/q1455/a-review-of-the-evidence-assessing-impact-of-social-prescribing-on-healthcare-demand-and-cost-implications [accessed 2021-02-18]
8. Longwill A. Independent Evaluation of Hackney Well Family Service. Improving HealthWellbeing UK. URL: https://www.family-action.org.uk/content/uploads/2014/07/WellFamily-Evaluation-Executive-Summary.pdf [accessed 2021-02-18]
9. How to introduce a social prescribing champion model and/or a digital platform to support your social prescribing offer. Healthy London Partnership. 2018. URL: https://www.healthylondon.org/wp-content/uploads/2018/12/How-to-introduce-a-social-prescribing-champion-model.pdf [accessed 2019-10-12]
10. Internet access- households and individuals, Great Britain. Office for National Statistics. 2019. URL: https://tinyurl.com/ykv2ph8t [accessed 2021-02-18]
11. Surge in the number of people using NHS tech during Coronavirus outbreak. 2020. URL: https://digital.nhs.uk/news-and-events/news/increase-in-nhs-tech-usage [accessed 2020-06-10]
12. Fogg BJ. Persuasive Computers: Perspectives and Research Directions. Persuasive Computers: Perspectives Research Directions. Proceedings of the CHI 98 Conference on Human Factors in Computing Systems; 1998 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 1998; Los Angeles. [doi: 10.1145/274644.274677]
13. Fogg B, Nass C. Silicon sycophants: the effects of computers that flatter. International Journal of Human-Computer Studies 1997 May;46(5):551-561 [FREE Full text] [doi: 10.1006/ijhc.1996.0104]
14. Evergreen Life and PCG launch personalised digital social prescribing solution. Evergreen Life. 2018. URL: https://www.evergreen-life.co.uk/news/evergreen-life-and-pcg-launch-personalised-digital-social-prescribing-solution [accessed 2018-12-10]
15. Elemental. URL: https://elementalsoftware.co [accessed 2019-04-04]
16. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
17. Pickton D, Wright S. What's swot in strategic analysis? Strat. Change 1998 Mar;7(2):101-109. [doi: 10.1002/(sici)1099-1697(199803/04)7:2<101::aid-jsc332>3.0.co;2-6]
18. Spiers J, Morse J, Olson K, Mayan M, Barrett M. Reflection/Commentary on a Past Article: "Verification Strategies for Establishing Reliability and Validity in Qualitative Research". International Journal of Qualitative Methods 2018 Aug 28;17(1):160940691878823 [FREE Full text] [doi: 10.1177/1609406918788237]
19. Mental Health Taskforce. The Five Year Forward View for Mental Health. 2016. URL: https://www.england.nhs.uk/wp-content/uploads/2016/02/Mental-Health-Taskforce-FYFV-final.pdf [accessed 2021-02-18]
20. Patel V, Maj M, Flisher AJ, De Silva MJ, Koschorke M, Prince M, WPA ZonalMember Society Representatives. Reducing the treatment gap for mental disorders: a WPA survey. World Psychiatry 2010 Oct;9(3):169-176 [FREE Full text] [doi: 10.1002/j.2051-5545.2010.tb00305.x] [Medline: 20975864]

21. Naslund J, Aschbrenner K, Araya R, Marsch L, Unützer J, Patel V, et al. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. The Lancet Psychiatry 2017 Jun;4(6):486-500 [FREE Full text] [doi: 10.1016/s2215-0366(17)30096-2]
22. Hampshire K, Porter G, Owusu SA, Mariwah S, Abane A, Robson E, et al. Informal m-health: How are young people using mobile phones to bridge healthcare gaps in Sub-Saharan Africa? Soc Sci Med 2015 Oct;142:90-99 [FREE Full text] [doi: 10.1016/j.socscimed.2015.07.033] [Medline: 26298645]
23. The digital future of mental healthcare and its workforce. Foley T. 2019. URL: https://topol.hee.nhs.uk/wp-content/uploads/HEE-Topol-Review-Mental-health-paper.pdf [accessed 2021-02-18]
24. Chatterjee H, Camic P, Lockyer B, Thomson L. Non-clinical community interventions: a systematised review of social prescribing schemes. Arts & Health 2017 Jun 06;10(2):97-123 [FREE Full text] [doi: 10.1080/17533015.2017.1334002]
25. Bickerdike L, Booth A, Wilson PM, Farley K, Wright K. Social prescribing: less rhetoric and more reality. A systematic review of the evidence. BMJ Open 2017 Apr 07;7(4):e013384 [FREE Full text] [doi: 10.1136/bmjopen-2016-013384] [Medline: 28389486]
26. Grant C, Goodenough T, Harvey I, Hine C. A randomised controlled trial and economic evaluation of a referrals facilitator between primary care and the voluntary sector. BMJ 2000 Feb 12;320(7232):419-423 [FREE Full text] [doi: 10.1136/bmj.320.7232.419] [Medline: 10669447]
27. Garrett S, Elley C, Rose S, O'Dea D, Lawton B, Dowell A. Are physical activity interventions in primary care and the community cost-effective? A systematic review of the evidence. Br J Gen Pract 2011 Mar 01;61(584):e125-e133 [FREE Full text] [doi: 10.3399/bjgp11x561249]
28. Pescheny, Pappas Y, Randhawa G. Facilitators and barriers of implementing and delivering social prescribing services: a systematic review. BMC Health Serv Res 2018 Feb 07;18(1):86 [FREE Full text] [doi: 10.1186/s12913-018-2893-4] [Medline: 29415720]
29. Husk K, Blockley K, Lovell R, Bethel A, Lang I, Byng R, et al. What approaches to social prescribing work, for whom, and in what circumstances? A realist review. Health Soc Care Community 2020 Mar;28(2):309-324 [FREE Full text] [doi: 10.1111/hsc.12839] [Medline: 31502314]
30. Hollis, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265 [FREE Full text] [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
31. Pavey, Taylor A, Hillsdon M, Fox K, Campbell J, Foster C, et al. Levels and predictors of exercise referral scheme uptake and adherence: a systematic review. J Epidemiol Community Health 2012 Aug;66(8):737-744 [FREE Full text] [doi: 10.1136/jech-2011-200354] [Medline: 22493474]
32. Pescheny J, Randhawa G, Pappas Y. Patient uptake and adherence to social prescribing: a qualitative study. BJGP Open 2018 Aug 07;2(3):bjgpopen18X101598. [doi: 10.3399/bjgpopen18x101598]
33. Mills H, Crone D, James V, Johnston L. Exploring the Perceptions of Success in an Exercise Referral Scheme. Eval Rev 2013 May 02;36(6):407-429. [doi: 10.1177/0193841x12474452]
34. Baker K, Irving A. Co-producing Approaches to the Management of Dementia through Social Prescribing. Social Policy & Administration 2015 Mar 16;50(3):379-397. [doi: 10.1111/spol.12127]
35. Stathi A, McKenna J, Fox K. The experiences of older people participating in exercise referral schemes. J R Soc Promot Health 2004 Jan;124(1):18-23. [doi: 10.1177/146642400312400108] [Medline: 14971188]
36. Jones F, Harris P, Waller H, Coggins A. Adherence to an exercise prescription scheme: the role of expectations, self-efficacy, stage of change and psychological well-being. Br J Health Psychol 2005 Sep;10(Pt 3):359-378. [doi: 10.1348/135910704X24798] [Medline: 16238853]
37. Kreps G. Stigma and the Reluctance to Address Mental Health Issues in Minority Communities. In: Journal of Family Strengths; 17(1): The TMC Library; 2017.
38. Landeiro F, Barrows P, Nuttall Musson E, Gray AM, Leal J. Reducing social isolation and loneliness in older people: a systematic review protocol. BMJ Open 2017 May 17;7(5):e013778 [FREE Full text] [doi: 10.1136/bmjopen-2016-013778] [Medline: 28515187]
39. Internet access- households and individuals, Great Britain. Office for National Statistics. URL: https://tinyurl.com/29jzuvsw [accessed 2021-02-18]
40. Schmidt EA, Wykes T. E-mental health - a land of unlimited possibilities. J Ment Health 2012 Aug;21(4):327-331 [FREE Full text] [doi: 10.3109/09638237.2012.705930] [Medline: 22823092]
41. Healthcare consumers in a digital transition. Rock Health. 2018. URL: https://rockhealth.com/reports/healthcare-consumers-in-a-digital-transition [accessed 2019-01-15]
42. Friedli L. Social prescribing for mental health – a guide to commissioning and delivery. Care Services Improvement Partnership North West Development Centre. URL: https://www.centreforwelfarereform.org/uploads/attachment/339/social-prescribing-for-mental-health.pdf [accessed 2019-10-15]
43. Rosales A, Fernández-Ardèvol M. Ageism in the era of digital platforms. Convergence 2020 Jun 25;26(5-6):1074-1087. [doi: 10.1177/1354856520930905]

44. Silva S, Kenney M. Algorithms, platforms, and ethnic bias. Commun. ACM 2019 Oct 24;62(11):37-39. [doi: 10.1145/3318157]
45. Fazel S, Singh JP, Doll H, Grann M. Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people: systematic review and meta-analysis. BMJ 2012 Jul 24;345:e4692 [FREE Full text] [doi: 10.1136/bmj.e4692] [Medline: 22833604]
46. Independent ROTMHAC. Modernising the mental health act: increasing choice, reducing compulsion. London. URL: https://www.gov.uk/government/groups/independent-review-of-the-mental-health-act [accessed 2018-11-06]
47. The Data Ethics Canvas. Open Data Institute. 2017. URL: https://theodi.org/article/data-ethics-canvas [accessed 2021-02-18]
48. Apter C. The benefits of social prescribing- and a word of warning. Mental health today. URL: https://www.mentalhealthtoday.co.uk/news/awareness/the-benefits-of-social-prescribing-and-a-word-of-warning [accessed 2019-09-25]

Abbreviations

DSPT: digital social prescription tool
GP: general practitioner
SWOT: strengths, weaknesses, opportunities, and threats

Edited by G Eysenbach; submitted 12.12.19; peer-reviewed by N Jacobson, M Mulvenna, G Dosovitsky; comments to author 16.04.20; revised version received 27.07.20; accepted 30.01.21; published 09.03.21.

Please cite as:
Patel S, Craigen G, Pinto da Costa M, Inkster B
Opportunities and Challenges for Digital Social Prescribing in Mental Health: Questionnaire Study
J Med Internet Res 2021;23(3):e17438
URL: https://www.jmir.org/2021/3/e17438
doi:10.2196/17438
PMID:33687338

Original Paper

Informatics Methodology Used in the Web-Based Portal of the NASCITA Cohort Study: Development and Implementation Study

Michele Zanetti[1], BEng; Antonio Clavenna[1], MD; Chiara Pandolfini[1], PhD; Claudia Pansieri[1], PharmD; Maria Grazia Calati[1], HS; Massimo Cartabia[1], StatSc; Daniela Miglio[1], HS; Maurizio Bonati[1], MD

Laboratory for Mother and Child Health, Department of Public Health, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy

Corresponding Author:
Chiara Pandolfini, PhD
Laboratory for Mother and Child Health
Department of Public Health
Istituto di Ricerche Farmacologiche Mario Negri IRCCS
Via Mario Negri 2
Milan, 20156
Italy
Phone: 39 02 3901 4478
Email: chiara.pandolfini@marionegri.it

Abstract

Background: Many diseases occurring in adults can be pinned down to early childhood and birth cohorts are the optimal means to study this connection. Birth cohorts have contributed to the understanding of many diseases and their risk factors.

Objective: To improve the knowledge of the health status of Italian children early on and how it is affected by social and health determinants, we set up a longitudinal, prospective, national-level, population-based birth cohort, the NASCITA study (NAscere e creSCere in ITAlia). The main aim of this cohort is to evaluate physical, cognitive, and psychological development; health status; and health resource use in the first 6 years of life in newborns, and potential associated factors. A web-based system was set up with the aim to host the cohort; provide ongoing information to pediatricians and to families; and facilitate accurate data input, monitoring, and analysis. This article describes the informatics methodology used to set up and maintain the NASCITA cohort with its web-based platform, and provides a general description of the data on children aged over 7 months.

Methods: Family pediatricians were contacted for participation in the cohort and enrolled newborns from April 2019 to July 2020 at their first well-child visit. Information collected included basic data that are part of those routinely collected by the family pediatricians, but also parental data, such as medical history, characteristics and lifestyle, and indoor and outdoor environment. A specific web portal for the NASCITA cohort study was developed and an electronic case report form for data input was created and tested. Interactive data charts, including growth curves, are being made available to pediatricians with their patients' data. Newsletters covering the current biomedical literature on child cohorts are periodically being put up for pediatricians, and, for parents, evidence-based information on common illnesses and problems in children.

Results: The entire cohort population consists of 5166 children, with 139 participating pediatricians, distributed throughout Italy. The number of children enrolled per pediatrician ranged from 1 to 100. The 5166 enrolled children represent 66.55% (5166/7763) of the children born in all of 2018 covered by the same pediatricians participating in the cohort. The number of children aged over 7 months at the time of these analyses, and for whom the most complete data were available upon initial analyses, was 4386 (2226/4381 males [50.81%] and 142/4370 twins [3.25%]). The age of the mothers at birth of the 4386 children ranged from 16 to 54 years. Most newborns' mothers (3758/4367, 86.05%) were born in Italy, followed by mothers born in Romania (101/4367, 2.31%), Albania (75/4367, 1.72%), and Morocco (60/4367, 1.37%). Concerning the newborns, 138/4386 (3.15%) were born with malformations and 352/4386 (8.03%) had a disease, most commonly neonatal respiratory distress syndrome (n=52), neonatal jaundice (n=46), and neonatal hypoglycemia (n=45).

Conclusions: The NASCITA cohort is well underway and the population size will permit significant conclusions to be drawn. The key role of pediatricians in obtaining clinical data directly, along with the national-level representativity, will make the findings even more solid. In addition to promoting accurate data input, the multiple functions of the web portal, with its interactive platform, help maintain a solid relationship with the pediatricians and keep parents informed and interested in participating.

Trial Registration: ClinicalTrials.gov NCT03894566; https://clinicaltrials.gov/ct2/show/NCT03894566

(J Med Internet Res 2021;23(3):e23087) doi:10.2196/23087

KEYWORDS

internet; computer systems; cohort studies; pediatricians; infant; newborn

Introduction

It is well known that many diseases occurring in adults can be traced back to early childhood [1,2]. In fact, nearly all domains of later health experience, including cardiovascular and respiratory disease, cognitive decline, and psychological health, have been associated with early life exposures [3]. Many different factors in childhood play a role in future health inequalities between individuals, from socioeconomic status to parental care, to lifestyle factors, but the way they are related is uncertain.

Birth cohort studies are studies that follow a group of newborns for an extended period in order to assess possible associations between exposures in early life and later health. Northern Europe has a long-lasting tradition in birth cohorts [4,5], starting from as far back as 1921 [6]. Findings from these studies have led to important knowledge in different fields, contributing to the understanding of multiple diseases and their risk factors [7,8]. These studies have also set the basis of our positive daily health behaviors. The Avon Longitudinal Study of Parents and Children (ALSPAC), for example, showed that eating oily fish during pregnancy was associated with better eye and cognitive development in children [9].

Numerous large- and small-scale birth cohorts have been set up, also in the past decade, not only in Europe but all around the world [10]. Characteristics vary greatly from one cohort to another in terms of design, objectives, size, and duration of follow-up.

Since 2003, several cohorts have also been carried out in Italy. Most of them have general aims, with data collection limited in time or to specific geographical contexts [11-19]. Italy is a special country with a public, universal health care system that should be equally accessible to all, but considerable health inequalities exist [20,21]. Up to now, no national-level birth cohort has been set up that included a large sample of the pediatric population independent of socioeconomic status or other types of limitation, such as gestational age. The Piccolipiù cohort [17], for example, recruited newborns from northern and central Italy; the NINFEA cohort [16] population was limited to women who had enough knowledge of the internet to complete online questionnaires; and the ICON cohort [19] selected preterm newborns and enrolled additional newborns of later gestational age.

In order to improve the knowledge of the health status of Italian children early on and how it is affected by social and health determinants, we set up a longitudinal, prospective, national-level, population-based birth cohort, the NASCITA study (NAscere e creSCere in ITAlia) [22]. Like many other cohorts, it addresses multiple research questions [16,17]. NASCITA is unique, however, in terms of characteristics, methodology, and population size. The findings will add important evidence, in terms of epidemiological data, for the development of specific prevention measures and interventions to improve the health status of children.

The main aim of the NASCITA cohort is to evaluate physical, cognitive, and psychological development, and health status and health resource use during the first 6 years of life in a group of newborns, and to evaluate potential associated factors.

The peculiarity of NASCITA is that data collection is designated to the general pediatricians, fitting itself into the Italian public health care system, as data reported in NASCITA are part of those routinely collected by the family pediatricians at the well-child visits. Furthermore, the data are equally distributed throughout the Italian territory.

A website and web-based system [23] were set up in order to host the cohort, provide ongoing information to pediatricians and to families, and facilitate data input on the part of the pediatricians. The system was also designed to optimize data accuracy, minimize missing data, and permit data monitoring, analysis, and reporting throughout the duration of the cohort.

This article describes the informatics methodology used to set up and maintain the NASCITA cohort with its web-based platform, and provides a general description of the participant characteristics.

Methods

Cohort Organization

NASCITA is embedded in Italian pediatric primary care practice. Data collection for the NASCITA cohort occurs for the most part during the 7 well-child visits planned for each child. The majority of the participating pediatricians are part of the national Pediatric Cultural Association (ACP), an association with about 2000 members consisting mainly of family pediatricians and with which the coordinating center has collaborated over the years. Participation was proposed to the ACP and forms the basis of pediatrician participation in the study. Pediatrician participation was voluntary and for free. Collaboration was also expanded through contact with other pediatric scientific societies and associations. Meetings were held during 2018 to present the study to a group of pediatricians acting as local representatives. Each representative then asked other pediatricians working in their area (at the local health unit or regional levels) to participate. Pediatrician enrollment was monitored and discussed with the local representatives. A scientific committee was set up to supervise the project, and includes professionals and lay people from different fields of expertise.

At the start of the study, there were 7960 cities/towns in Italy. These were classified into 21 geographic clusters (Figure 1), identified based on geographic and administrative criteria used by the Italian National Statistics Institute (ISTAT) [24]. More specifically, these take into consideration geographic area (north, center, south), setting (urban, rural), and land characteristics

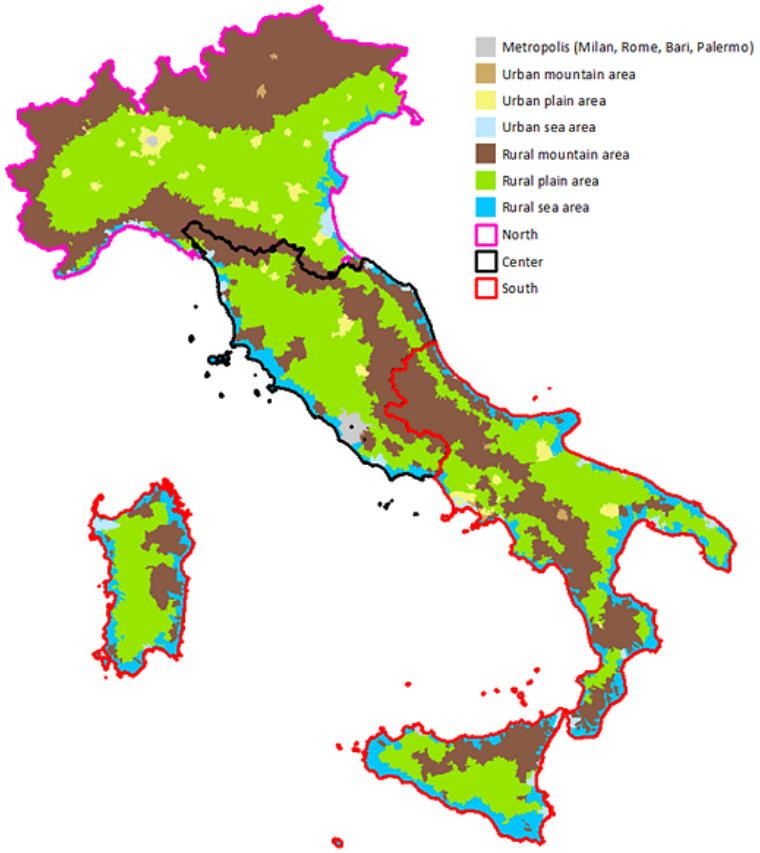
(plain, mountain, sea). Four cities were also selected (Milan, Rome, Bari, and Palermo), covering the different geographic areas and the islands.

Enrollment of newborns began in April 2019 and ended in July 2020. Recruitment of the newborns (and their parents) took place during the first routine well-child visit scheduled for all newborns in Italy within their first 45 days of life. All newborns assigned to the participating pediatricians were enrolled if parental consent was given. Pediatricians chose when to begin enrolling their newborns and continued to enroll for (at least) a 1-year period. Follow-up of the children will continue until at least the age of 6 years.

A minimum recruitment of 5000 newborns was calculated in order to have enough power to study common childhood exposures and outcomes.

In this article we present the characteristics of the children aged over 7 months at the time of analyses in order to provide more complete data, as pediatricians would have had time to fill in most missing data for these participants.

Figure 1. The 21 geographic clusters, identified based on ISTAT geographic and administrative criteria. ISTAT: Italian National Statistics Institute.



Ethics

Parents were given oral and written information about the study and a consent form to sign if interested in participating. Pediatricians signed a consent form before participation as well. Withdrawal from the study was guaranteed at any time to both pediatricians and parents.

Data Collection

Italian health care is provided free or at a nominal charge through a network of 148 local health units. The local health units assign children to a family pediatrician until they are 6 years old, after which the children can be registered with a general practitioner or remain with that pediatrician until they are 14 years old. In Italy there are about 7500 family pediatricians, for an average of 450,000 births/year [25], so about 60 newborns/year are assigned to each pediatrician. All children are scheduled 7 well-child visits at the pediatrician's office during their first 6 years of life to ensure necessary preventive care and monitor a child's growth and development.

Basic data are being collected and consist of data that are part of those routinely collected by the family pediatricians at the well-child visits. Some data will also be collected during each extra contact with the enrolled children. Data collection also involves parental data, such as medical history, characteristics and lifestyle, indoor and outdoor environment, and circumstances during pregnancy and around birth. Follow-up data on children will cover different fields, including physical and mental/cognitive development, nutrition and allergies, environmental exposures, and preventable infectious diseases. See Table 1 for a description of the main parts of the questionnaire. Questions were added to allow the project, in a second phase, to address specific areas such as nutrition, environment, and nurturing care.

Table 1. Main sections of the online questionnaires and description of the general data collected.

Section	Description
Personal data	Name, place of birth, family data such as number of family members, sibling health, parental place of birth, allergies
Medical history	Mother's pregnancy data (including medicines, smoking and alcohol consumption, and reading out loud and listening to music), birth data (eg, newborn height, weight), perinatal medical history (eg, malformations, diseases, transfer to an intensive care unit), breastfeeding status at discharge
Visits 1-7	Medicines taken, anthropometric measures; breastfeeding status/weaning/nutrition; sleep data; age-appropriate physical examination; vitamin D + K prophylaxis; psychomotor, neurologic, and cognitive development; general health; paternal depression; language development; family habits (eg, smoking, reading out loud, listening to music, nursery school, indoor and open-air activities); home proximity to traffic or to areas of intensive farming; screen time
Extra visits	Type of contact (office, phone, home visit), diagnosis, medicines/specialist visits/examinations prescribed
Vaccination compliance	Vaccines received and adverse reactions
Exiting the cohort	Reason for exit (transfer, pulled out of study, death)

Statistical Analysis

The analyses of the cohort data will evaluate specific research questions related to the overall aims of the study, such as the relationship between child development and the domains that affect nurturing care during the preschool period including health, nutrition, and caregiving routine; the association between the well-being of children and parental adherence to the recommendations for better child care and development; and the differences between geographical settings in educational and socialization opportunities available and in the care provided by the family pediatricians.

Data are presented as frequencies, percentages, and mean (SD) or median values. Percentages are based on denominators for which missing values have been excluded. All data management and analyses have been performed using SAS version 9.4 (SAS Institute) and ArcMap version 10.5 (ESRI). More detailed analyses will be performed, as specified in the protocol [22], and reported in future articles.

Web Portal

A specific web portal for the NASCITA cohort study was developed [23] with reserved sections for the coordinating center, registered users, and participating pediatricians. The web portal serves to assist pediatricians with data collection and to provide findings and other information during the study period to parents and pediatricians, also with the use of graphics for the analyses and data collected, based on a successful approach already reported by the coordinating center [26]. Selected sections of the portal have been translated into English. See Figure 2 for the functions of the web portal and its architecture.

Newsletters focused on child cohorts are periodically provided in the pediatrician's general area and contain bibliographic information of the current biomedical literature. In the private area, each participating pediatrician can access information including cohort documents; frequently asked questions; the study protocol; and pdf versions of the case report forms (CRFs); as well as patient data for input/modification; interactive data charts of his/her patients or of those of the entire cohort, including growth curves (Figure 3); and data concerning subsections of the cohort addressing areas such as nutrition and environment in which he/she participates. The pediatrician's section on the web portal, together with individual telephone calls with the pediatricians and online and in-person meetings on the study's progress and possible problems, serves to keep pediatricians engaged in the study.

The information for the parents section contains a growing series of cards, created in collaboration between health professionals and parents, that provide evidence-based information on the more common illnesses and problems in young children as well as answers to common questions that parents have on child care. This section also contains links to useful emergency telephone numbers and information pages.

Figure 2. Functions of the platform and its architecture.

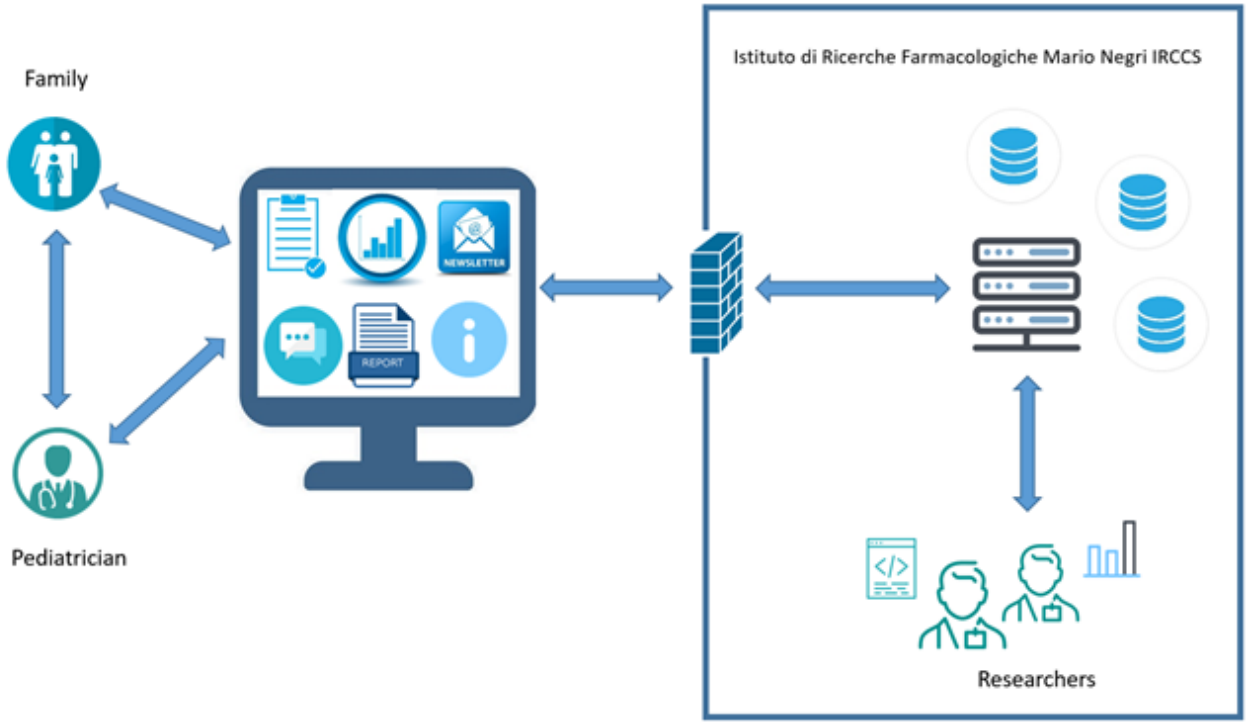
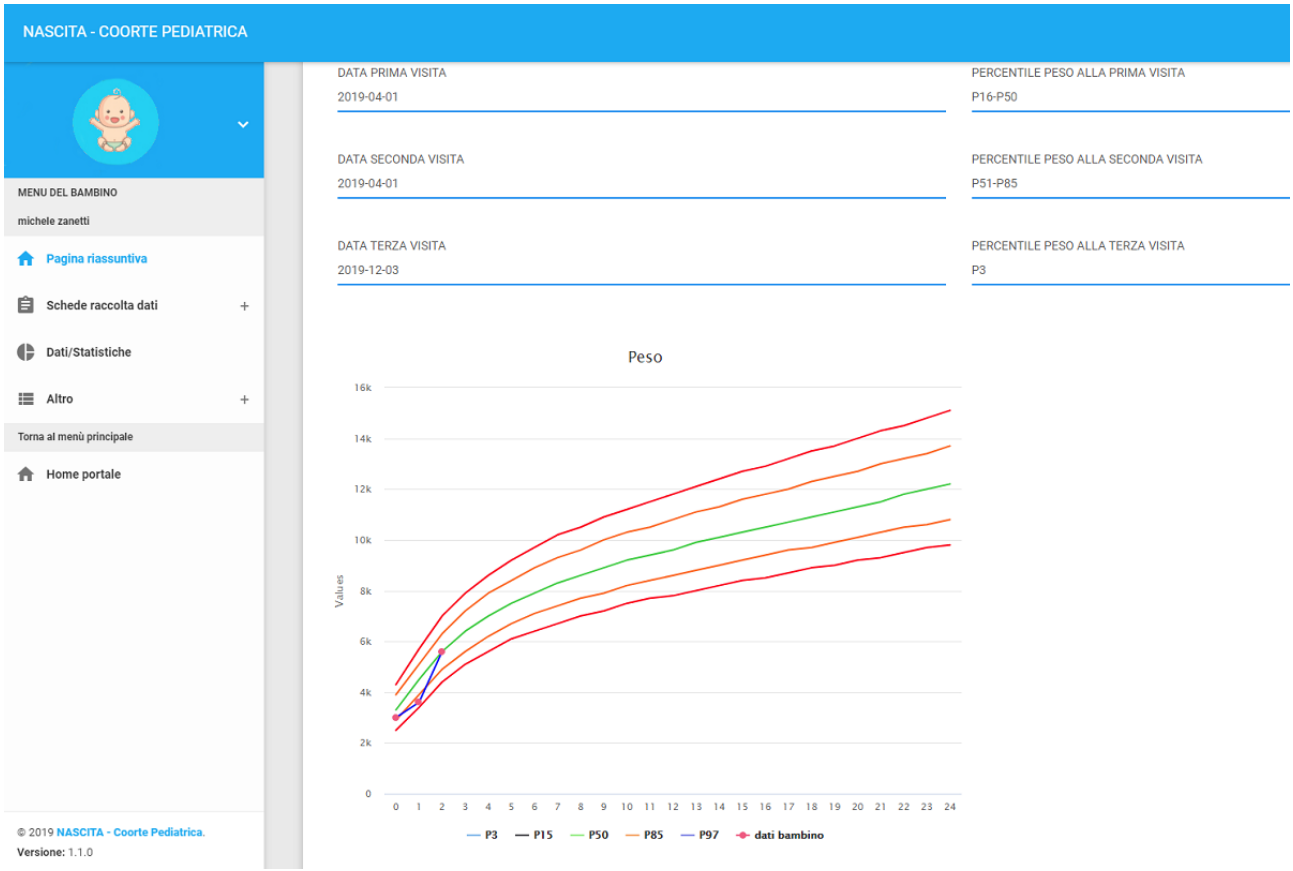


Figure 3. Individual child's data plotted on growth curve, as seen by the pediatrician.



Electronic Case Report Form

The CRF was created and tested together with a group of pediatricians. An electronic CRF (eCRF) was then set up and tested, again with the help of the pediatricians, before enrollment began. More specifically, the pediatricians were asked to register themselves and access the portal starting from January 2019 to test it. The eCRF (Figure 4) was set up in such a way as to facilitate the pediatricians' input of data for the study and to provide fast and efficient support for any problems or doubts about data input. A "chat" section was consequently included through which pediatricians can ask for support. The eCRF

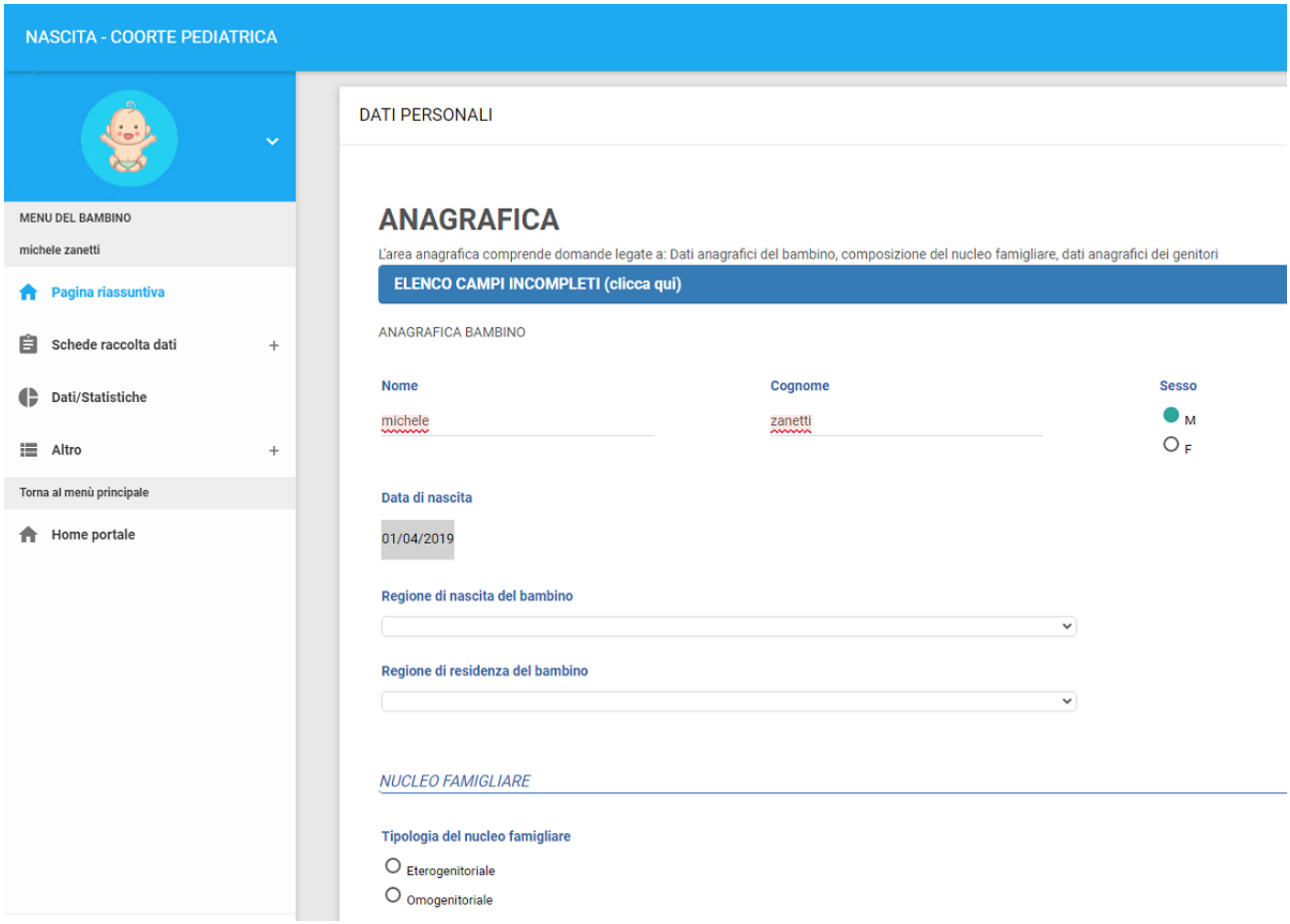
includes consistency and range checks to prevent internal inconsistencies (eg, value ranges, fields with limited values, and time ranges). Data are, in any case, monitored continuously and irregularities resolved through email, chat, or phone contact with the pediatricians.

The eCRF has been structured in a way that will permit data collection to be expanded to cover the additional areas (eg, nutrition) more thoroughly in a second phase; the different data collection sections are, in fact, based on an XML definition that can easily be implemented and modified [27].

The development engine of the eCRF has been made available on Gitlab [28].

A test was performed with a group of pediatricians (including those with less experience in using the computer) to assess the additional amount of time it would take each pediatrician to enter data for a patient throughout the duration of the study. Entering data for the first follow-up visit took about 15 minutes. Multiplying this by the average number of newborns per pediatricians and, considering that after 3 months of the start of the study the subsequent regular check-ups would begin, an average of 3 hours a month in the first year was calculated, after which the amount of time necessary would decrease.

Figure 4. eCRF screenshots: input for second visit.

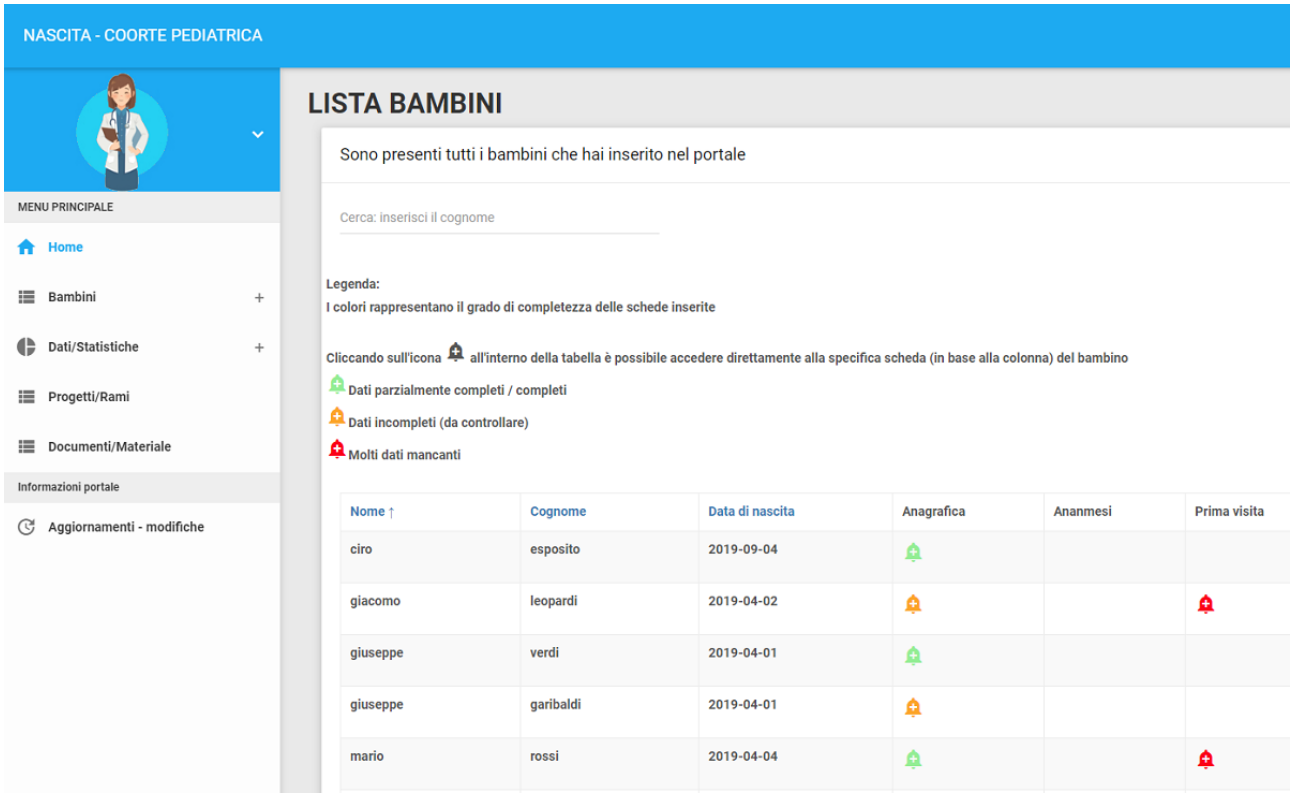


Data Quality Control

A dashboard is dedicated to checking the completeness, or lack, of the visits and displays a table listing the children for whom data have been included by that pediatrician. Each column in the table represents a specific visit and shows a series of colored bells (green, yellow, or red) that indicate the completeness status of the visit (Figure 5). When a pediatrician opens the data on a specific visit, the system displays a list of the variables with missing information in order to facilitate data completion.

Frequent reports will automatically be created to monitor recruitment of pediatricians and children and the data inputted. Individual and group reports will also be created for the pediatricians and for the study's scientific committee.

Figure 5. eCRF dashboard for checking completeness of data on the visits for individual patients.



Data Security

The private area is accessible only through authentication by pediatricians, who have been previously approved and enabled. A specific role is assigned to them, for example, *compiler*. Passwords must have at least eight characters and contain special and uppercase characters. Each pediatrician with a compiler role can insert their patients' data. Sensitive data, such as name and last name, are encoded and visible only to the compiler. Once the data are saved, they are transmitted via secure HTTPS protocol, and are stored in databases that can be accessed only by authorized project staff (IT, statisticians). Back-ups are kept for security and disaster recovery.

Results

Enrollment in the NASCITA cohort began on April 1, 2019, and ended on July 31, 2020. The number of participating pediatricians is 139 and the total number of children enrolled is 5166. The pediatricians are distributed throughout Italy, with 68 in the north, 29 in the center, and 42 in the south. The 5166 children enrolled represent 66.55% (5166/7763) of the children born in all of 2018 covered by these same pediatricians.

The total number of children aged over 7 months at the time of these analyses was 4386. Of these children, excluding those with missing data, 2226/4381 (50.81%) were male and 142/4370 (3.25%) were twins. The children were distributed throughout Italy, with 2025/4386 (46.17%) in the north, 882/4386 (20.11%) in the center, and 1479/4386 (33.72%) in the south. Table 2 reports the distribution of the number of these 4386 children enrolled and the percentage of children born, by cluster and geographic area, based on ISTAT data [24], and shows that there are minimal differences. The number of children enrolled per pediatrician ranged from 1 to 100 (mean 32 [SD 18.5]; median 30).

The age of the mothers at birth of the 4386 children ranged from 16 to 54 (mean 33 [SD 5.4] years; median 33 years, excluding 133 missing values), while the age of the fathers ranged from 17 to 69 (mean 36 [SD 6.3]; median 36, excluding 154 missing values). Most of the newborn's mothers (3758/4367, 86.05%) were born in Italy; the 3 next most common countries were Romania (101/4367, 2.31%), Albania (75/4367, 1.72%), and Morocco (60/4367, 1.37%). For two-thirds of children (2892/4320, 66.94%), the mothers were married or living in civil union; for 1233/4320 children (28.54%), the mother was living with the father; and for 167/4320 (3.87%), mothers were single. Most of the children's mothers had a university (1813/4320, 41.96%) or high-school degree (1800/4320, 41.66%), followed by a middle-school diploma (675/4320, 15.63%) and an elementary (30/4320, 0.69%) level of education. Family size (including the enrolled child) was grouped into 2, 3, 4, or >4 people, with half (2189/4311, 50.78%) of the families being made up of 3 people, followed by 1567/4311 (36.35%) made up of 4 people. Two-member families represented 1.43% (62/4311) of the total. Concerning the pregnancies, 3741/4363 (85.74%) were physiologic pregnancies, while in the remaining pregnancies gestational diabetes (203/622 mothers, 32.64%), gestational hypertension (90/622, 14.47%), and preeclampsia (41/622, 6.59%) were the most common diseases. Concerning the newborns, 139/4340 (3.20%) were born with malformations and 352/4335 (8.12%) had a disease, the 3 most common of which were neonatal respiratory distress syndrome (n=52), neonatal jaundice (n=46), and neonatal hypoglycemia (n=45).

Table 2. Distribution of the number of children and the number of children born in each cluster and geographic area.

Distribution/Location	Children enrolled, n (N=4386)	Population enrolled, n (%)	Births per year in Italy[a], n (%)	Difference between the third and fourth columns, %
Metropolis: Milan	184	184/4386 (4.20)	11,267/467,640 (2.41)	1.79
Metropolis: Rome	428	428/4386 (9.76)	21,497/467,640 (4.60)	5.16
Metropolis: Bari	70	70/4386 (1.60)	2214/467,640 (0.47)	1.13
Metropolis: Palermo	62	62/4386 (1.41)	5578/467,640 (1.19)	0.22
North: urban mountain	—[b]	—	2243/467,640 (0.48)	–0.48
North: urban plain	304	304/4386 (6.93)	45,595/467,640 (9.75)	–2.82
North: urban sea	93	93/4386 (2.12)	10,258/467,640 (2.19)	0.07
North: rural mountain	228	228/4386 (5.20)	25,274/467,640 (5.40)	–0.20
North: rural plain	1151	1151/4386 (26.24)	112,824/467,640 (24.13)	2.11
North: rural sea	60	60/4386 (1.37)	5882/467,640 (1.26)	0.11
Center: urban plain	61	61/4386 (1.39)	6330/467,640 (1.35)	–0.04
Center: urban sea	1	1/4386 (0.02)	5339/467,640 (1.14)	–1.12
Center: rural mountain	90	90/4386 (2.05)	13,887/467,640 (2.97)	–0.92
Center: rural plain	233	233/4386 (5.31)	31,997/467,640 (6.84)	–1.53
Center: rural sea	68	68/4386 (1.55)	11,118/467,640 (2.38)	–0.83
South: urban mountain	8	8/4386 (0.18)	4010/467,640 (0.86)	–0.68
South: urban plain	136	136/4386 (3.10)	15,826/467,640 (3.38)	–0.28
South: urban sea	421	421/4386 (9.60)	27,281/467,640 (5.83)	3.77
South: rural mountain	98	98/4386 (2.23)	21,932/467,640 (4.69)	–2.46
South: rural plain	421	421/4386 (9.60)	47,026/467,640 (10.06)	–0.46
South: rural sea	255	255/4386 (5.81)	40,262/467,640 (8.61)	–2.80
Missing	14	—	—	—

[a]Based on data on the newborn population residing in the cluster on January 1, 2017. ISTAT demographic statistics data [24] referring to December 31, 2016, were used for clusters with missing data.

[b]—: not available.

Discussion

Considerations

The NASCITA cohort is based on community-level pediatric practice, involving the family pediatricians directly, as very few European cohorts do [29]. With their clinical practice, pediatricians are most in contact with patients and can promote study and action. Their involvement in child cohorts permits the collection of prospective, community-level data and allows them to contribute to optimizing both the quality of the data collected and its re-investment back into the community as health promotion interventions. In fact, pediatricians play a key role both in educating families and in implementing curative and disease prevention interventions through their routine clinical practice. They are in the optimal position to influence public health in general because adult health also depends on habits embraced when young, and pediatricians can undoubtedly influence children and their parents to adopt healthy lifestyles. In order to give something back to the pediatricians participating in the cohort, we have attempted to provide the pediatricians with useful information and interesting data, such as the interactive data charts of their patients. The system set up through which pediatricians can easily and quickly contact the cohort team for any questions or problems, and the periodic meetings organized to update pediatricians on the cohort's status and to discuss any current issues or suggestions are additional ways to show our appreciation for all their continuing efforts. During the latest meeting we had with the participating pediatricians, online in November 2020, we described the enrolled population as it was just after enrollment closure and the next steps. On this occasion several pediatricians provided additional suggestions for improving input, resulting in the message to all that their participation and efforts are ongoing and continue to be acknowledged by the research team.

Recruitment of newborns took place over a period of 1 year for each pediatrician. This time span permitted us to avoid introducing bias related to the period of recruitment, for example, by recruiting newborns born during one season as opposed to another. The sample of children aged over 7 months reflects the distribution of births in Italy in terms of both geographic area (north, center, and south) and 21 clusters, based on the ISTAT data. Collection of data at the national level will

permit the identification of differences in health care quality, such as those caused by socioeconomic inequalities present between the north and south of Italy [30], and of differences in family behaviors that influence child health status (eg, smoking or reading out loud to children). Better identifying health care–related inequalities will permit the channeling of resources where they are most needed [31]. If funding is obtained, the population enrolled could be expanded further.

As explained previously, the web portal has multiple functions and is fundamental for several reasons. The innovative aspects involve permitting the accurate input and monitoring of data through the use of a tool that creates data collection based on an xml definition, and providing pediatricians with interactive charts of current data to share with the children's parents.

This xml-based system allows a continuous and simple updating of the CRF, saving a lot of time in the development and testing phase. In addition, saving the data in the JavaScript Object Notation (JSON) format allows greater flexibility in the database structure which, therefore, does not need to be remodified at each CRF update [32].

Furthermore, our idea for the future is to interconnect the portal with apps for parents to use to access data and to provide additional information.

Strengths and Limitations

This is one of a very limited number of child cohorts based on the participation of family pediatricians, permitting the collection of data by those directly involved with the children and the implementation of findings to inform and help those directly involved (the children and their families). Furthermore, the large, representative population sample of newborns throughout the country, which allows stratified trends based on socioeconomic and geographic characteristics to be performed, and the use of standard measurements for anthropometric and neurocognitive parameters are among the strengths of this study.

A limit of the NASCITA cohort is that it does not collect biological samples due to the costs of data collection and storage, so it will not be able to evaluate genetic or immunological factors, for example. Resources and efforts were utilized, however, to achieve the largest population size possible in order to have enough power to study relatively common child exposures and outcomes. Another limitation is the potential bias in the pediatrician population because participation was voluntary and this may have led more motivated pediatricians to participate than others.

Conclusions

The NASCITA cohort is well underway and its population size will permit significant conclusions to be drawn. The key role of pediatricians in obtaining clinical data directly, along with the national-level representativity, will make the findings even more solid. In addition to promoting accurate data, the multiple functions of the web portal, with its expanding, interactive platform, will help maintain a solid relationship with the pediatricians and keep parents informed and interested in participating.

Acknowledgments

The authors thank all the participating pediatricians and committee members for their work and dedication. We also thank Davide Bazzi and Massimo Vitali for assistance in creating and in the support of servers, databases, and back-up system. No direct funding was received. The costs were covered by resources from the Laboratory for Mother and Child Health and by an economic contribution by the Associazione Amici del Mario Negri and by 2 anonymous donors, all of whom we wish to thank.

Authors' Contributions

MB, AC, and C Pansieri conceptualized the study, designed the web portal, and designed the CRF. AC, C Pandolfini, C Pansieri, and DM tested the CRF. MZ created the web portal, the eCRFs, and the interactive charts system. AC, MC, and C Pansieri coordinated and monitored data collection. MGC collaborated in assisting the pediatricians with data input problems and managed email and phone communication with pediatricians and with the technical and scientific committees. MC carried out the analyses. DM and C Pansieri were responsible for merging requests for modifications and corrections to the CRFs during testing, for re-testing them, and for creating the layout with data check and limit specifications for transformation into the electronic version. DM was responsible for website content and created the information cards for parents and the newsletters for the pediatricians. C Pandolfini wrote the first draft of the manuscript. All authors interpreted data and critically reviewed and revised the final manuscript as submitted. All authors agree to be accountable for aspects of their contribution to the work. MB is the guarantor.

Conflicts of Interest

None declared.

References

1. Black MM, Walker SP, Fernald LCH, Andersen CT, DiGirolamo AM, Lu C, Lancet Early Childhood Development Series Steering Committee. Early childhood development coming of age: science through the life course. Lancet 2017 Jan 07;389(10064):77-90 [FREE Full text] [doi: 10.1016/S0140-6736(16)31389-7] [Medline: 27717614]

2. Barouki R, Gluckman PD, Grandjean P, Hanson M, Heindel JJ. Developmental origins of non-communicable disease: implications for research and public health. Environ Health 2012 Jun 27;11:42 [FREE Full text] [doi: 10.1186/1476-069X-11-42] [Medline: 22715989]
3. Fleming TP, Watkins AJ, Velazquez MA, Mathers JC, Prentice AM, Stephenson J, et al. Origins of lifetime health around the time of conception: causes and consequences. Lancet 2018 May 05;391(10132):1842-1852 [FREE Full text] [doi: 10.1016/S0140-6736(18)30312-X] [Medline: 29673874]
4. Larsen PS, Kamper-Jørgensen M, Adamson A, Barros H, Bonde JP, Brescianini S, et al. Pregnancy and birth cohort resources in europe: a large opportunity for aetiological child health research. Paediatr Perinat Epidemiol 2013 Jul;27(4):393-414. [doi: 10.1111/ppe.12060] [Medline: 23772942]
5. Thompson L, Kemp J, Wilson P, Pritchett R, Minnis H, Toms-Whittle L, et al. What have birth cohort studies asked about genetic, pre- and perinatal exposures and child and adolescent onset mental health outcomes? A systematic review. Eur Child Adolesc Psychiatry 2010 Jan;19(1):1-15 [FREE Full text] [doi: 10.1007/s00787-009-0045-4] [Medline: 19636604]
6. Baxter-Jones AD, Cardy AH, Helms PJ, Phillips DO, Smith WC. Influence of socioeconomic conditions on growth in infancy: the 1921 Aberdeen birth cohort. Arch Dis Child 1999 Jul;81(1):5-9 [FREE Full text] [doi: 10.1136/adc.81.1.5] [Medline: 10373124]
7. Framingham Heart Study. URL: https://framinghamheartstudy.org/ [accessed 2021-02-16]
8. Nurses' Health Study. URL: https://www.nurseshealthstudy.org/ [accessed 2021-02-16]
9. Pearson H. Children of the 90s: Coming of age. Nature 2012 Apr 11;484(7393):155-158. [doi: 10.1038/484155a] [Medline: 22498607]
10. Birthcohorts.net. URL: https://www.birthcohorts.net/ [accessed 2021-02-16]
11. Porta D, Fantini M. Prospective cohort studies of newborns in Italy to evaluate the role of environmental and genetic characteristics on common childhood disorders. Ital J Pediatr 2006 Dec;32:350-357.
12. Porta D, Forastiere F, Di Lallo D, Perucci CA, Grupo Collaborativo GASPII. [Enrolment and follow-up of a birth cohort in Rome]. Epidemiol Prev 2007;31(6):303-308 [FREE Full text] [Medline: 18326421]
13. Lanari M, Vandini S, Adorni F, Prinelli F, Di Santo S, Silvestri M, Study Group of Italian Society of Neonatology on Risk Factors for RSV Hospitalization. Prenatal tobacco smoke exposure increases hospitalizations for bronchiolitis in infants. Respir Res 2015 Dec 22;16:152 [FREE Full text] [doi: 10.1186/s12931-015-0312-5] [Medline: 26695759]
14. Barchitta M, Maugeri A, La Rosa MC, Magnano San Lio R, Favara G, Panella M, et al. Single Nucleotide Polymorphisms in Vitamin D Receptor Gene Affect Birth Weight and the Risk of Preterm Birth: Results From the. Nutrients 2018 Aug 27;10(9):1172 [FREE Full text] [doi: 10.3390/nu10091172] [Medline: 30150529]
15. Brescianini S, Fagnani C, Toccaceli V, Medda E, Nisticò L, D'Ippolito C, et al. An update on the Italian Twin Register: advances in cohort recruitment, project building and network development. Twin Res Hum Genet 2013 Feb;16(1):190-196. [doi: 10.1017/thg.2012.85] [Medline: 23088847]
16. Richiardi L, Baussano I, Vizzini L, Douwes J, Pearce N, Merletti F. Feasibility of recruiting a birth cohort through the Internet: the experience of the NINFEA cohort. Eur J Epidemiol 2007;22(12):831-837. [doi: 10.1007/s10654-007-9194-2] [Medline: 17955333]
17. Farchi S, Forastiere F, Vecchi Brumatti L, Alviti S, Arnofi A, Bernardini T, et al. Piccolipiù, a multicenter birth cohort in Italy: protocol of the study. BMC Pediatr 2014 Feb 07;14:36 [FREE Full text] [doi: 10.1186/1471-2431-14-36] [Medline: 24506846]
18. Valent F, Mariuz M, Bin M, Little D, Mazej D, Tognin V, et al. Associations of prenatal mercury exposure from maternal fish consumption and polyunsaturated fatty acids with child neurodevelopment: a prospective cohort study in Italy. J Epidemiol 2013 Sep 05;23(5):360-370 [FREE Full text] [doi: 10.2188/jea.je20120168] [Medline: 23933621]
19. Lanari M, Prinelli F, Adorni F, Di Santo S, Vandini S, Silvestri M, Study Group of Italian Society of Neonatology on Risk Factors for RSV Hospitalization. Risk factors for bronchiolitis hospitalization during the first year of life in a multicenter Italian birth cohort. Ital J Pediatr 2015 May 26;41:40 [FREE Full text] [doi: 10.1186/s13052-015-0149-z] [Medline: 26006025]
20. Bonati M, Campi R. What can we do to improve child health in southern Italy? PLoS Med 2005 Sep;2(9):e250 [FREE Full text] [doi: 10.1371/journal.pmed.0020250] [Medline: 16104833]
21. Landi S, Ivaldi E, Testi A. Socioeconomic status and waiting times for health services: An international literature review and evidence from the Italian National Health System. Health Policy 2018 Apr;122(4):334-351. [doi: 10.1016/j.healthpol.2018.01.003] [Medline: 29373188]
22. Pansieri C, Clavenna A, Pandolfini C, Zanetti M, Calati MG, Miglio D, et al. NASCITA Italian birth cohort study: a study protocol. BMC Pediatr 2020 Feb 19;20(1):80 [FREE Full text] [doi: 10.1186/s12887-020-1961-1] [Medline: 32075601]
23. Nascere e crescere in Italia (Coorte NASCITA). URL: http://nascita.marionegri.it/ [accessed 2021-02-16]
24. Istituto Italiano di Statistica (ISTAT). Main geographic statistics on cities. URL: https//www.istat.it/it/archivio/156224 [accessed 2021-02-16]
25. Corsello G, Ferrara P, Chiamenti G, Nigri L, Campanozzi A, Pettoello-Mantovani M. The Child Health Care System in Italy. J Pediatr 2016 Oct;177S:S116-S126. [doi: 10.1016/j.jpeds.2016.04.048] [Medline: 27666260]

26. Bonati M, Reale L, Zanetti M, Cartabia M, Fortinguerra F, Capovilla G, Lombardy ADHD Group. A Regional ADHD Center-Based Network Project for the Diagnosis and Treatment of Children and Adolescents With ADHD. J Atten Disord 2018 Oct;22(12):1173-1184. [doi: 10.1177/1087054715599573] [Medline: 26320120]
27. Zanetti M, Campi R, Olivieri P, Campiotti M, Faggianelli A, Bonati M. A Web-Based Form With Interactive Charts Used to Collect and Analyze Data on Home Births in Italy. J Med Internet Res 2019 Mar 22;21(3):e10335 [FREE Full text] [doi: 10.2196/10335] [Medline: 30900993]
28. GitLab. Gatherings Crafter. URL: https://gitlab.com/dellaltromondo/gatherings-crafter [accessed 2020-07-20]
29. Pansieri C, Pandolfini C, Clavenna A, Choonara I, Bonati M. An Inventory of European Birth Cohorts. Int J Environ Res Public Health 2020 Apr 28;17(9):3071 [FREE Full text] [doi: 10.3390/ijerph17093071] [Medline: 32354101]
30. de Waure C, Bruno S, Furia G, Di Sciullo L, Carovillano S, Specchia ML, et al. Health inequalities: an analysis of hospitalizations with respect to migrant status, gender and geographical area. BMC Int Health Hum Rights 2015 Feb 07;15:2 [FREE Full text] [doi: 10.1186/s12914-014-0032-9] [Medline: 25881154]
31. Lauria L, Spinelli A, Buoncristiano M, Nardone P. Decline of childhood overweight and obesity in Italy from 2008 to 2016: results from 5 rounds of the population-based surveillance system. BMC Public Health 2019 May 21;19(1):618 [FREE Full text] [doi: 10.1186/s12889-019-6946-3] [Medline: 31113403]
32. JavaScript Object Notation. URL: https://www.w3.org/TR/html-json-forms/#the-application-json-encoding-algorithm [accessed 2021-02-16]

Abbreviations

ACP: Pediatric Cultural Association
CRF: case report form
eCRF: electronic case report form
ISTAT: National Statistics Institute
JSON: JavaScript Object Notation
NASCITA cohort: NAscere e creSCere in ITAlia cohort

Edited by G Eysenbach; submitted 31.07.20; peer-reviewed by F Adorni, J Dol; comments to author 18.09.20; revised version received 12.11.20; accepted 18.01.21; published 12.03.21.

Please cite as:
Zanetti M, Clavenna A, Pandolfini C, Pansieri C, Calati MG, Cartabia M, Miglio D, Bonati M
Informatics Methodology Used in the Web-Based Portal of the NASCITA Cohort Study: Development and Implementation Study
J Med Internet Res 2021;23(3):e23087
URL: https://www.jmir.org/2021/3/e23087
doi:10.2196/23087
PMID:33709930

Original Paper

A Semiautomated Classification System for Producing Service Directories in Social and Health Care (DESDE-AND): Maturity Assessment Study

Cristina Romero-Lopez-Alberca[1,2], PhD; Federico Alonso-Trujillo[3,4], MD, PhD; Jose Luis Almenara-Abellan[4,5], MD; Jose A Salinas-Perez[6,7], PhD; Mencia R Gutierrez-Colosia[8], PhD; Juan-Luis Gonzalez-Caballero[9], PhD; Sandra Pinzon Pulido[10], MD, PhD; Luis Salvador-Carulla[7], MD, PhD

[1]Department of Psychology, Universidad de Cádiz, Cádiz, Spain

[2]Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain

[3]Agencia de Servicios Sociales y Dependencia de Andalucía, Junta de Andalucía, Sevilla, Spain

[4]Health Information Systems Group (SICA-CTS-553), Universidad de Cádiz, Cádiz, Spain

[5]Hospital Universitario Reina Sofía, Servicio Andaluz de Salud, Córdoba, Spain

[6]Department of Quantitative Methods, Universidad Loyola Andalucía, Sevilla, Spain

[7]Centre for Mental Health Research, Research School of Population Health, ANU College of Health and Medicine, Australian National University, Canberra, Australia

[8]Department of Psychology, Universidad Loyola Andalucía, Sevilla, Spain

[9]Departamento Estadística e Investigación Operativa, Universidad de Cádiz, Cádiz, Spain

[10]Escuela Andaluza de Salud Pública, Gobierno Regional de la Junta de Andalucía, Granada, Spain

Corresponding Author:
Federico Alonso-Trujillo, MD, PhD
Agencia de Servicios Sociales y Dependencia de Andalucía
Junta de Andalucía
C/ Seda 5 - Polígono Hytasa
Sevilla, 41071
Spain
Phone: 34 609548503
Email: federico.alonso@juntadeandalucia.es

Abstract

Background: DESDE-LTC (Description and Evaluation of Services and DirectoriEs for Long-Term Care) is an international classification system that allows standardized coding and comparisons between different territories and care sectors, such as health and social care, in defined geographic areas. We adapted DESDE-LTC into a computer tool (DESDE-AND) for compiling a directory of care services in Andalucia, Spain.

Objective: The aim of this study was to evaluate the maturity of DESDE-AND. A secondary objective of this study is to show the practicality of a new combined set of standard evaluation tools for measuring the maturity of health technology products.

Methods: A system for semiautomated coding of service provision has been co-designed. A panel of 23 domain experts and a group of 68 end users participated in its maturity assessment that included its technology readiness level (TRL), usability, validity, adoption (Adoption Impact Ladder [AIL]), and overall degree of maturity [implementation maturity model [IMM]). We piloted the prototype in an urban environment (Seville, Spain).

Results: The prototype was demonstrated in an operational environment (TRL 7). Sixty-eight different care services were coded, generating fact sheets for each service and its geolocation map. The observed agreement was 90%, with moderate reliability. The tool was partially adopted by the regional government of Andalucia (Spain), reaching a level 5 in adoption (AIL) and a level 4 in maturity (IMM) and is ready for full implementation.

Conclusions: DESDE-AND is a usable and manageable system for coding and compiling service directories and it can be used as a core module of decision support systems to guide planning in complex cross-sectoral areas such as combined social and health care.

(J Med Internet Res 2021;23(3):e24930) doi:10.2196/24930

KEYWORDS

DESDE-LTC; DESDE-AND; services coding; service directories; decision support system; impact analysis; maturity

Introduction

COVID-19 has dramatically exposed the cracks of otherwise highly connected systems such as social and health services. For example, nursing homes are both health-related organizations and social housing facilities for highly vulnerable people [1]. This connection was mostly hidden to service planners until the patients' and staff flows between hospitals and nursing homes during COVID-19 unleashed "a perfect storm" around the health and social care systems locally and globally [2]. This pandemic has also shown the relevance of computer modeling, care navigation systems, real-time dashboards, and interactive decision support systems to guide evidence-informed policy. However, to be effective, rapid response digital systems should go beyond traditional semantic interoperability across data sets, and move toward a common coding and counting of service provision (availability, bed capacity, and workforce) from a whole system's perspective (health, social, housing, employment, education, and justice) [3-5]. This harmonization effort should follow a service ecosystem approach [6] and focus on the comprehensive assessment of the overall local and regional service availability and capacity as it has also been underscored by the pandemic [1]. The World Health Organization (WHO) Advisory Committee on Health Research has pointed out to the importance of developing useful policies based on evidence of the availability of services in catchment areas [7,8]. This cross-sectorial coding system of service provision should also allow international comparison of social and health systems, as this is essential to detect possible gaps in service capacity and provision and to conduct comparative effectiveness studies.

However, the aggregation and comparison of services from a whole system's perspective is not an easy task, as shown by previous developments in disability or functional dependency in Europe [9,10], older adults [11], and mental health and chronic care [12]. These attempts have been recently extended to education [3] and justice [4] where similar problems have been identified.

The difficulty of comparability between services across sectors is related to a number of factors including [13-15] (1) the geographic variability of care systems; (2) the noncommensurability bias or lack of consensus on the units of analysis for comparisons at macro (countries, regions), meso (small health areas), and micro (individual services) levels [16]; (3) the terminological variability of the services and programs, and the lack of a workable formal ontology of service provision in the main health terminology systems such as SNOMED; (4) and the low usability and comparability of official directories or listings of services available at the regional or national level. These factors highlight the importance of a consensus on the taxonomy and classification of services analogous to that existing for the classification of diseases, classification of operations, and classification of health interventions (eg, in the World Health Organization Family of International Classifications [17]); and the importance of using instruments for service assessment with published psychometric data that may facilitate disambiguation and enhance semantic interoperability in service assessment and monitoring.

DESDE-LTC (Description and Evaluation of Services and DirectoriEs for Long-Term Care) is a standardized coding system for care services to be used across sectors and tools such as health and social directories, atlases of care, and decision support systems [10,12,18,19]. This system identifies and codes basic care teams (named Basic Stable Inputs of Care [BSIC]) as its main unit of analysis and applies a code to their main activity called "Main Type of Care." It provides local bottom–up multisectoral coding [20] of care services across different target groups and has been tested in mental health care [21,22], intellectual disability [23], alcohol and drug abuse [24,25], general disability [10], aging, and chronic or long-term care [16,26]. DESDE-LTC has been translated into 9 languages and applied in over 34 countries [27]. The information gathered using this classification system has proven its usefulness for designing decision support systems to guide regional mental health planning in Spain [28] and Australia [29] when combined with local information on the context of care (eg, sociodemographic data), the use of resources (eg, activities, interventions), and indicators of results [28,30].

Despite its uniqueness and tested psychometric properties, the applicability of DESDE-LTC is limited due to the significant effort in data gathering that requires surveys and direct interviews to managers of local services, as well as an intensive training for coders [15]. Mapping an urban health or social district (eg, between 300,000 and 1 million inhabitants) takes over 6 months and is a research-intensive task [31]. Therefore, an automated version of this ontology system is needed for its routinization in local and regional care planning.

This study evaluates the maturity of DESDE-AND, an online computer system to produce a semiautomatic classification of services for compiling standard directories of care based on DESDE-LTC. "Maturity" has been defined in software development as the potential capability of the process of implementation of a new product and the consistency with which it could be applied in projects throughout the target organization [32]. The capability maturity model is used in business and technology sectors for establishing life cycle and planning sustainment and has been applied to the evaluation of the process of implementation of medical products and in health care research [33]. A secondary objective of this study is to show the practicality of a new combined set of standard evaluation tools for measuring the maturity of health technology products.

Methods

Study Design

This study follows a co-design/hybrid approach. The co-design process was based on information provided by a domain expert panel and key end users. The initial expert panel consisted of 28 professionals from different social and health areas of the regional public Agency for Social Services and Dependency of Andalucia (ASSDA). This group included experts in information technologies, policy planning, management, and service research. Training on the DESDE-LTC coding and on the use of the DESDE-AND prototype (version 0.0) was provided to all the experts. From this group, 23 domain experts participated in the panel discussion of the maturity study. The group of end users included the managers of the 68 services that participated in the demonstration study.

Setting

The tool has been designed in a collaboration partnership between Psicost Research Association and the public ASSDA. Within the regional Ministry of Equity and Social Policies of Andalucia (Spain), ASSDA coordinates the social public and private care and the welfare system for the 8.5 million inhabitants living in this Spanish region [34]. The regional government has full governance on planning, funding, and managing its health and social care systems. Apart from social services, the Ministry of Equity and Social Policies regulates and manages a complex mix of health and social services for specific target groups (eg, drug and alcohol, aging, disabilities) as well as other support services such as housing, special employment, and special education services. It also participates in the interministerial agency of mental health involving the departments of health and social services. The Social Services Law [35] has made mandatory the development of the Social Services Map of Andalucia, and recommends the implementation of a "classification system for the different types of services for planning and evaluation." The tool has been tested in the city of Seville (target area), with nearly 700,000 inhabitants.

Instruments

The DESDE-LTC is a tool developed from the European Service Mapping Schedule [36,37]. This instrument presents a hierarchical tree structure with 6 main service branches, at 6 levels of granularity, and its last version includes 124 possible codes to describe the typology of each service assessed. The tool is distributed into 4 sections. Section A includes the general principles of evaluation and coding, the description of the reference area, and the target population of the services to be codified in the assessment process. Section B consists of a tree that represents the hierarchical structure, the description of the codes, and their identification by means of labels of the main health care types. Section C collects the use of services (Basic Stable Inputs of Care) in a reference area. Finally, section D includes a service inventory that collects data on the main characteristics of each service.

Implementation Maturity Model (IMM)

The implementation maturity model (IMM) uses the 5 maturity levels from the capability maturity model to assess and determine the degree of maturity of implementation processes in software engineering [32,38]. In Level 1 (Initial), the organizational capability process is not structured, and is undefined and inconsistent; it largely relies on successes of individuals. In Level 2 (Repeatable), the process has partially consistent, successful processes; and Level 3 (Defined) has standard and documented processes. In Level 4 (Managed), the process is quantitatively managed, while the process is Optimized in Level 5 [33]. From these definitions, the level of maturity of the different aspects of the implementation process can be confirmed, thus describing their strengths and weaknesses, and also where improvements are needed. This ordinal rating is designed for self-assessment of the process of development of a product.

Technology Readiness Level (TRL)

Readiness is the level of preparedness for the application of a new scientific knowledge for commercialization or generalized use in the real world. The technology readiness levels (TRLs) are a systematic measurement that supports assessments of the maturity of a particular technology during the acquisition phase of a program. Nine levels are considered [39]: TRL 1, Basic principles observed and reported; TRL 2, Technology concept and/or application formulated; TRL 3, Analytical and experimental critical function and/or characteristic proof of concept; TRL 4, Component and/or breadboard validation in laboratory environment; TRL 5, Component and/or breadboard validation in relevant environment; TRL 6, System/subsystem model or prototype demonstration in a relevant environment; TRL 7, System prototype demonstration in a space environment; TRL 8, Actual system completed and "flight qualified" through test and demonstration; TRL 9, Actual system "flight proven" through successful mission operations.

Psicost Usability Checklist

The Usability Checklist consists of 15 items that assess relevance (meaningfulness, novelty, and potentiality of the new scientific knowledge to the target audience), acceptability (the degree to which a new application or product is agreed or approved by the target audience), functionality, security, practicality (related to implementation, training requirements, and complexity of evaluation, and the analysis, interpretation, and reporting of the data), efficiency (practicality in relation to effort, time, and costs), training, and agreement. It uses a 10-point Likert scale, in which 0 is none, and 10 is the maximum score of each assessed variable. The questionnaire can be adapted to the characteristics of any new tool and the target end users. It has been used in the assessment of decision support systems [30] and was adapted for the assessment of the DESDE-AND in the target organization [15]. The usability questionnaire allows the study of the feasibility of a tool in its initial version, as well as the analysis of its final usability after its pilot study.

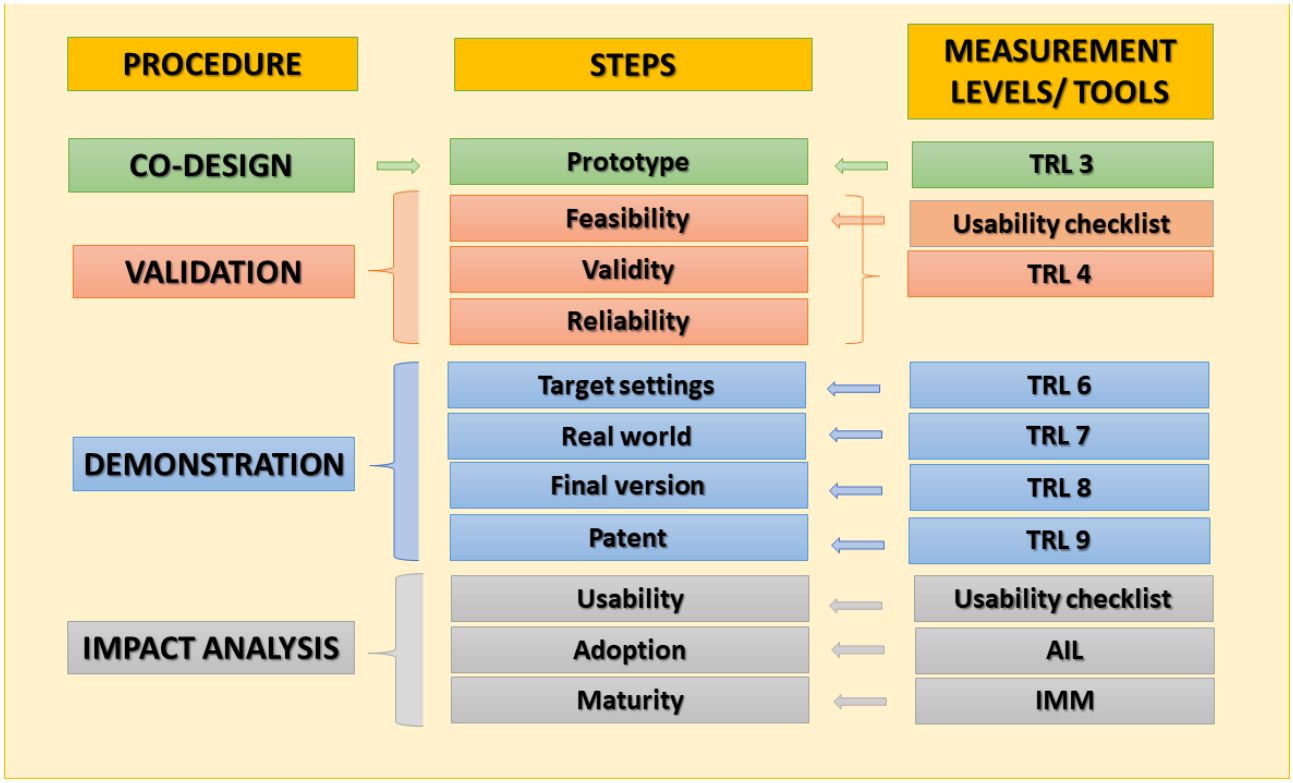
Adoption Impact Ladder (AIL)

The Adoption Impact Ladder (AIL) is an inventory for evaluating the level to which the target organization has taken the new knowledge or its application as its own. It uses a quasi-ordinal scale with 7 categories, namely, (0) no adoption; (1) awareness; (2) assimilation; (3) conversion (or translation); (4) allocation; (5) provision; and (6) routinization (or monitoring). It has been previously used in Australia for assessing the impact on the practice of a state policy program (Ed-LinQ Program in Queensland) [40], the international dissemination of a classification of case management [41], the adoption of the Spanish Disability Scheme by all the regional agencies in Andalucia [42], and in the use of DESDE-LTC for mapping mental health services in Spain [19]. Its validation study in Spain has indicated high usability and good reliability for assessing the impact of the adoption of multisectoral social and health policies and plans [42].

Procedure

This study consists of 4 stages comprising 11 steps and 4 measurement instruments (Figure 1).

Figure 1. Stages of the DESDE-AND study methodology. AIL: Adoption Impact Ladder; IMM: implementation maturity model; TRL: technology readiness level.



Stage 1: Co-design of the DESDE-AND Prototype

The tool was built as an online computer system that guides "end users" step by step. The end user of this computer tool was defined as the manager or other decision maker of a service depending on the regional Ministry of Equality, Social Policies and Conciliation of Andalucia.

The development of the computer tool used architectural standards and was specifically programmed using JAVA. The software developer company (Guadaltech S.L.) prepared the user/administrator manuals, installation guides, implementation manual, and source code.

Two additional computer modules were developed to explore the combined use of DESDE-AND as a core module of a DSS (a data export system linked to geographic information system [GIS] for visualization and spatial analysis) and to assess its usability.

Stage 2: Validation of the DESDE-AND Prototype (Version 0.0)

The validation process of the DESDE-AND was part of level 4 of the technology readiness. A mini-Delphi panel of 23 domain experts produced the expert's knowledge base [43] on the DESDE-AND prototype. During the proof-of-concept phase, the experts conducted an assessment of the feasibility of the prototype (ie, the usability during the development process). The Psicost Usability Checklist was used to assess the feasibility prototype after coding a set of case examples from the real world. We calculated the averages of the scores of the different domains of feasibility: relevance, acceptability, functionality, security, practicality, efficiency, and agreement. The interrater reliability was analyzed during the demonstration phase. Kappa coefficients were calculated by comparing the coding made by an expert in DESDE-LTC (alpha evaluator) and the 68 managers of the services selected for the demonstration phase (beta evaluators). Finally, we analyzed the predictive validity of the codes obtained using DESDE-AND.

Stage 3: Demonstration Study—DESDE-AND Version 0.1

This stage included a pilot study and the assessment of the DESDE-AND practicability and manageability. The pilot study was carried out in the Seville metropolitan area. A total of 188 services providing care authorized from the regional Ministry of Equality, Social Policies and Conciliation of Andalucia were identified in the target area and invited to participate in the study. This list included 13 services located outside the target area that provided types of care that were not available in the city of Seville but that could be used by their residents. As much as 73 services accepted to participate but 5 of them were not able to complete the information requested. Finally, 68 services participated in the demonstration study.

In order to perform the coding, the managers of the services (end users) identified the basic social and health care units of each service and automatically assigned them a "main type of care" code according to the decision tree system offered. This code was subsequently confirmed by the personnel responsible for public administration.

The panel of experts carried out the analysis of the results of the coding in the selected services, which allowed the preparation of a report with recommendations for improvement of the instrument.

The recommendations by the panel of experts allowed modifications and improvements to be made for the development of the instrument in its final version (version 1.0).

Stage 4: Impact Analysis

The Psicost Usability Checklist, which was first used to assess the feasibility of the prototype by the 23 domain experts, was completed at the end of the demonstration phase by the 68 end users to assess the usability of the tool. The adoption by the target organization was measured by 3 experts in service evaluation and in the use of the decision support tools (FA-T, MG-C, and JS-P). The level of adoption was assessed in the individual services and in the target organization using AIL.

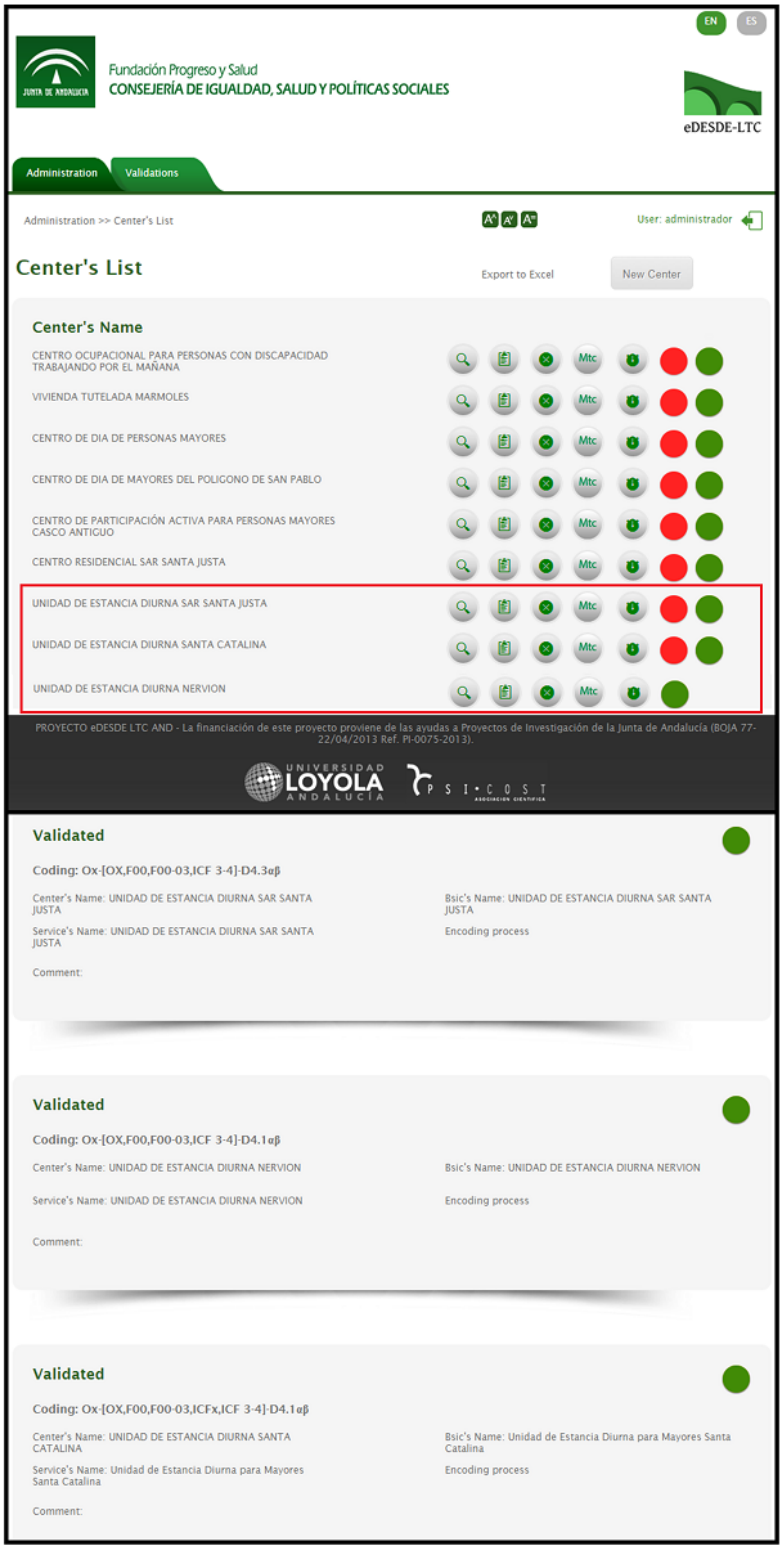
Every stage of the development process of the tool corresponded to a TRL. At the completion of the final version (DESDE-AND version 1.0), the group of domain experts assessed the level of technological maturity reached by the tool (TRL), which was compared with the one provided by the core team. Finally, the overall level of maturity was assessed using the IMM.

Results

Stage 1: Co-Design of the DESDE-AND Prototype

The schemas and algorithms necessary for the development of the computer-administered version 0.0 (prototype) were designed in the first stage. The classification process used specific question answering algorithms with 2 or more logical answers in a decision tree. Subsequently, a user interface for the web application was created with a user-friendly design, using a set of images and graphic objects to represent the available actions (Figure 2).

Figure 2. User interface of DESDE-AND web application.



A modular system with minimal interconnection between modules was used to develop the software and user interfaces.

The modular operability of DESDE-AND was tested with a supplementary module designed for exporting georeferenced information from the coding generated in the service inventory to any GIS and online cartographic viewers. This has been used to demonstrate the capacity of DESDE-AND data to be exported to other modules within a composite decision support system. GIS is the most common visualization system in DSS [44]. It allows performing complex spatial analyses, such as context studies (sociodemographic situation of the geographic environment of the services), assessing the optimal location of services and their accessibility. The results of the pilot study were integrated into the open GIS of Google Maps.

Stage 2: Validation of the DESDE-AND Prototype (Version 0.0)

The feasibility of the tool was assessed using the Usability Checklist. A total of 23 domain experts completed the instrument (Figure 3). The "Relevance" items were rated high (average score of 9). The "Acceptability" was relatively low, with an average score of 6.57. Questions about "Functionality" obtained an average score of 7.87. The items that assessed the "Practicality" of the tool obtained an average rating of 7.3, while "Efficiency" reached an average score of 7.57. Regarding the need for prior training, 80% (19/23) of the experts considered that the 68 service managers should have prior training at an advanced level, and 57% (13/23) of the experts considered that the advance training level should be also provided to the personnel that classified the services. The experts had an average agreement of 76.1%.

Figure 3. Experts' responses to the usability questionnaire. Each cell represents the qualitative score or answer of a different expert, and the last column shows the mean of their quantitative score by question and domain.

ITEM	SCORE	MEAN
RELEVANCE		
Q1. Relevance of standardized service classification systems for service planning and management. 0 = No importance – 10 = Maximum importance	High Very high	9.00
ACCEPTABILITY		
Q2. Assessment of the degree of resistance of the professionals who are in charge of coding the services of a center with the tool eDESDE-AND. 0 = Maximum resistance – 10 = No resistance	Very low Low High Very high	6.57
FUNCTIONALITY		
Q3. Usefulness of the DESDE-AND computer tool for coding social services. 0 = No usefulness – 10 = Maximum usefulness	Low High Very high	7.87
SECURITY		8.15
Q4. Information security guarantee. 0 = No security – 10 = Maximum security	Low High Very high	8.13
Q5. Protection mechanisms necessary to guarantee the security of the information from the DESDE-AND tool. 0 = No protection mechanism – 10 = Every protection mechanisms	Low High Very high	8.17
PRACTICALITY		7.30
Q6. Ease of use for coding social services. 0 = Not easy – 10 = Totally easy	Low High Very high	7.22
Q7. Ease in learning. 0 = Not easy – 10 = Totally easy	Low High Very high	7.65
Q8. Clarity and ease of interpretation of information to aid coding. 0 = Not clear – 10 = Totally clear	Low High Very high	6.78
Q9. Sufficiency of the documentation provided and help tabs for the cataloging of social services. 0 = Totally insufficient – 10 = Totally enough	Low High Very high	6.83
Q10. Availability of technical help during the testing process. 0 = No availability – 10 = Maximum availability	Low High Very high	8.04
EFFICIENCY		
Q11. Time spent for coding social services is reasonable. 0 = No reasonable – 10 = Totally reasonable	High Very high	7.57
TRAINING		
Q12. Need for training prior to the use of the tool by the administration professional who manages/supports the coding of services. No; User; Advanced; Expert	User Advanced Expert	
Q13. Need for training prior to the use of the tool by the administration professional who performs the coding of the services. No; User; Advanced; Expert	User Advanced Expert	
AGREEMENT		
Q14. Do you consider that there are clearly different services from each other that can be classified with the same code? Yes; No; N/A	N/A No Yes	
Q15. Do you consider that there are clearly similar services to each other that can be classified with different codes? Yes; No; N/A	N/A No Yes	

Scores: Very low: 0-2; Low: 3-4; High: 5-7; Very high: 8-10 N/A: Not Available

The interrater reliability analysis was performed after the pilot study in the 68 services. We assessed the coding concordance between ratings provided by the 68 service managers and the experts. The agreement was 0.89 with a Kappa index of 0.33 (95% CI). We also calculated the predictive validity of the tool (sensitivity, specificity, and predictive values), and observed high results in specificity (89.39%; 95% CI 81.21-97.58), sensitivity (100%), and predictive value (100%).

Stage 3: Demonstration Study—DESDE-AND Version 0.1

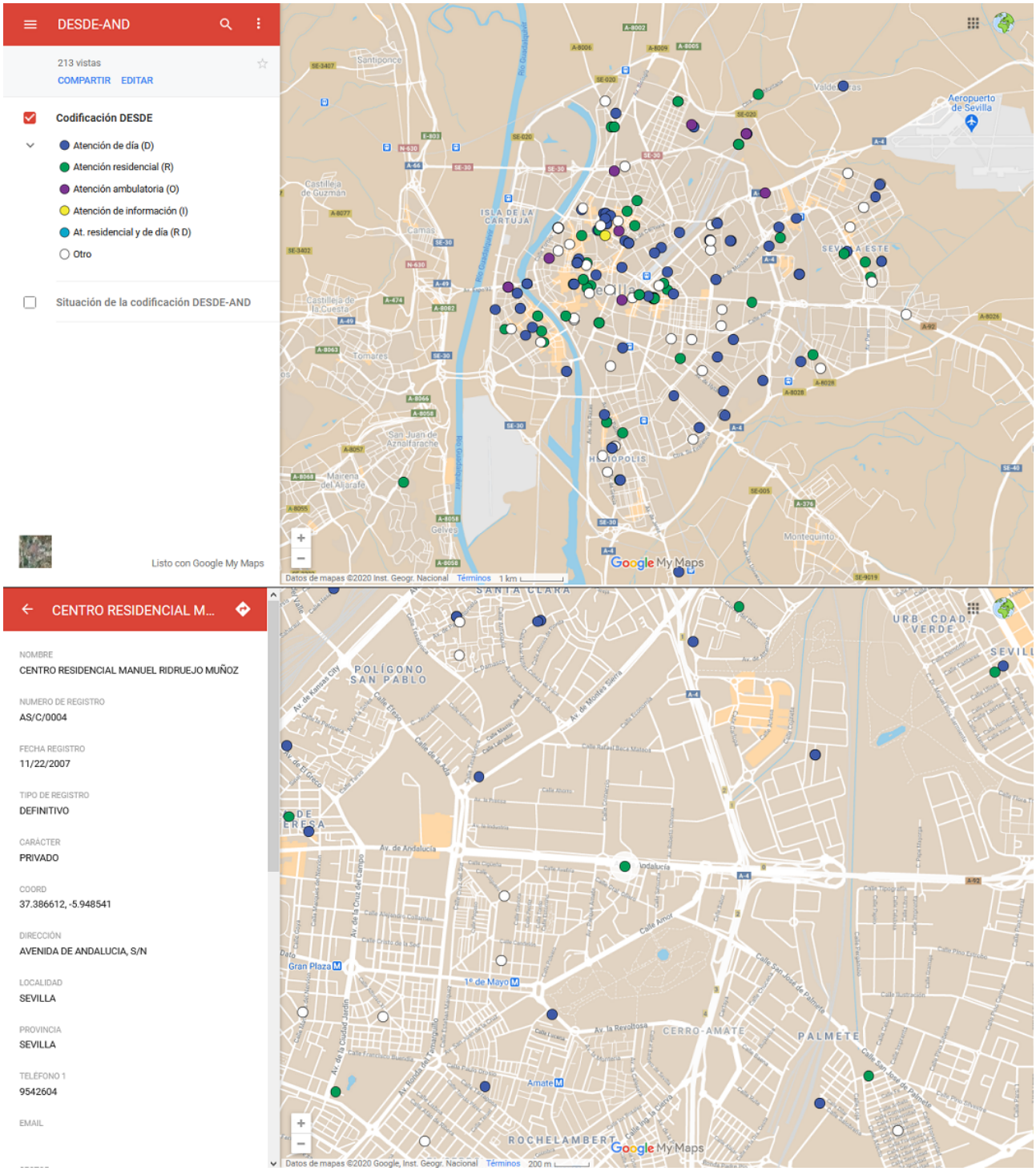
The pilot study included 68 services that agreed to participate and were classified using the DESDE-AND. The remaining services did not respond to the request, because they were closed or because they did not have adequate human or technical resources. Six of them started the process but did not complete the coding because they did not have personnel with basic training in information technology and communication. The 68 coded services were designed for different target groups: child and adolescents (n=5, 7%), older adults (n=40, 59%), persons with disabilities (n=16, 24%), drug addiction problems (n=5, 7%), and community social services (n=2, 3%).

The assessment of usability of the final version of DESDE-AND was performed by the service managers for the 68 services of the pilot study. Average scores were 7.6 for relevance, 7.8 for functionality, 7.8 for usability, 6.9 for efficiency, and 2.4 for resistance of the staff to perform the coding. More than half of the of the professionals (13/23, 57%) who coded the services considered that training was necessary at the user level in order to use the DESDE-AND.

We received 63 improvement suggestions and comments on the prototype to produce version 0.1. Of these, 52 recommendations were accepted and, among them, the corrections considered were performed to ensure gender equality. Eleven proposals were discarded due to technical issues, programming, or errors in the use of the tool.

The results of the pilot study in Seville were integrated into the open GIS of Google Maps [45], allowing the geolocation of the results of DESDE in any standard GIS system in the market. Figure 4 provides a depiction of the actual digital tool: the online directory of services, the classification using the DESDE coding system, and the geolocation of the services in the catchment area. This figure shows the importance of disambiguation in social care before they are geolocated. There are services that are officially identified with the same name that receive 2 different DESDE codes and are identified as different type in the GIS.

Figure 4. Screenshots of the web-mapping application (Google Maps) for the geographic visualisation of the services coded through DESDE-AND tool for social care planning in Andalucia (Spain).



Stage 4: Impact Analysis

The level of adoption was assessed in the individual services registered for the study as well as in the main target organization (ASSDA) by 3 experts (FA-T, MRG-C, and JAS-P). According to these experts, DESDE-AND reached an AIL level of 5 (impact on the budget, financing, and allocation of resources for testing the tool in the target environment without routine use within the organization).

The overall maturity level of the tool was moderate-high or Managed (IMM 4): the process of implementing is documented throughout the organization rather than per aspect, and projects are carried out under the guidance of project operation standards and an implementation strategy [38].

This process has generated a computer tool with a semiautomated core module to obtain the standard directory of the health and social services available in a given area, based on an international classification.

Discussion

Principal Findings

It uses a multiaxial system that allows typify entities into differentiated groups using a string of codes or "axes" corresponding to the main attributes of the entity from other groupings in the same system [46]. This multiaxial system classifies the services according to the geographical catchment area, the target population, and the type of care provided (residential care, day care, outpatient care, information, accessibility, or self-help). It includes a file with detailed information relating to each of the available services reducing disambiguation and enhancing the semantic interoperability of health and social care databases. The improvement of the applicability and practicality ratings of DESDE-AND in comparison with the ones obtained by the paper and pencil version [15] indicates a better usability, although the actual time for completion and updating the coding and mapping of care districts will require a further analysis after routinization.

This classification can be used by managers with the expert support provided by the computer tool, simplifying the complex task of coding in real time, and unifying the taxonomic and semantic criteria. It has been developed using a co-design approach focused on end users (managers, planners, and other decision makers) [47,48]. It ensures that the design satisfies the users of the system by making improvements in the usability and human performance, thus reducing previous training requirements in length and intensity.

DESDE-AND is a system that allows various functions and different levels of web surfing/navigation and use: (1) a simple easy navigation for consumers; (2) a second advanced level of web surfing for case managers to improve care continuity; and (3) an advanced level that can be incorporated to decision support tools for social and health planning. This tool facilitates the analysis of health care ecosystems [6], covering the various sectors, levels, and types of services involved in the care provision of an area, as well as the local characteristics and the drivers of the system (socioeconomic, demographic, legislative, and political). This information is relevant for recognizing and understanding the pattern of care provision, care gaps, unmet needs, and duplications of any system [20].

This system has been developed using a co-design approach. Co-design is largely used to improve service quality and delivery [49], and can also be used for the purpose of policy making [50,51]. Using co-design to inform policy making is particularly important to mental health care as we recognize that co-creation of well-being is critical and principal actors are vulnerable [48,52]. Our study uses co-design to improve the quality of indicators by accounting for contextual and environmental factors as well as to foster knowledge transfer of decision and policy making prior to policy design. By doing so, the system can afford to inform and bridge the resource gaps between the quality of service provision and the quality of decision and policy.

The original instrument DESDE-LTC has demonstrated its use in national and international comparisons of service directories and atlases of health and social care [27]. As DESDE-AND uses the same coding system, it will improve the comparisons across jurisdictions. This is key for service research, traditionally hampered by intrinsic problems in assessments of services [13]. This tool also facilitates the identification of complex patterns of provision and differences in national and international health and care research and monitoring [16,22,53]. It could improve the analysis of associations between the level of availability of services and sentinel indicators such as deprivation or social fragmentation [54], suicide rates [55], family burden [56], employment [57], or the cost of illnesses [58].

The information related to the coding of services can be exported to a GIS. This modular function facilitates data filtering by regions or provinces, municipalities, sectors of care, and types of services. It enables rapid geo-positioning of services with operational identification criteria and their specific characteristics. Henceforth, DESDE-AND facilitates the development of social and health care atlases of care [19]. Atlases are one of the visual tools currently being used for the assessment and planning of health systems [44].

The basic information obtained also allows the development of service indicators classified by geographic study areas, such as rates of services availability per inhabitants, beds (placement capacity), and professionals (workforce capacity); the analysis of the diversity of care in an area; and the balance between health and social care [30]. It could also be used to analyze accessibility and optimal locations of services. Atlases using the DESDE system have been made available in Spain [59], Europe [22,60], Australia [31,61], and Chile [53].

The final implementation of the tool in Andalusia depends on a series of regulatory steps. The Andalusian Social Services Law of 2017 entails the Map of Social Services of Andalusia and its update. The DESDE coding system will be included in the next version of the map.

The Map of Social Services of Andalucia is in operation. It is a computerized map of social services that allows the geolocation of the centers and services in the Catalogue of Social Services of Andalucia [62]. The inclusion of a field with an international reference code for the standardized classification of each center typology, such as DESDE-AND, constitutes an example of possible practical applications of the tool in planning social services in a certain area.

Finally, and as a secondary result, this study illustrates the use of standard measures and instruments for the quantitative assessment of maturity, a key phase in implementation research.

Limitations

First, this computer tool uses an internet-based algorithm adapted to the specific conditions of the Autonomous Community of Andalucia, and to the information data sets of the Andalusian Social Services. The transferability to other territories has not been tested. However, the study describes the process and tools required to produce similar semiautomated tools adapted to other environments. Second, there is a lack of international agreement and guidelines for conducting impact analysis of digital health care [63] and the domains of maturity that require evaluation. We opted for the overall ordinal

assessment of the maturity level (IMM [38]), complemented by the ordinal rating of key domains of maturity instead of conducting in-depth interviews to stakeholders. "Maturity" is considered here as part of the "early implementation" phase [64].

Although the maturity assessment instruments used in this study were developed for self-rating and monitoring within organizations, they can provide useful information on the calibration of impact analysis across different projects, areas, and sectors in health technology assessment. We used a mixed approach combining evidence and expert knowledge, and by applying a co-design process [65] selected a defined catchment area using a health care ecosystem approach [20,66,67], and identified the target organizations in this catchment area. The maturity general rating [33] and the maturity domains selected in this study are supported by previous literature in health care technology assessment: the "TRL" (adopted by the European Commission and Horizon programs) [39]; the "Usability" of the Tool in a real world environment [15]; and its "Adoption" by target organizations, as suggested by Health Quality Ontario [66].

Conclusions

DESDE-AND is a usable and manageable computer tool for guiding evidence-informed decision making in health and social planning. This tool reduces ambiguity and increases semantic interoperability in health and social care. It improves a previous tool that has demonstrated its usability for detecting gaps and inequities in the provision of health and social care. DESDE-AND is a relevant contribution to establish a common terminology, classification, and coding of health and social care services in the national context, as well as a standardized procedure for data collections and comparisons.

Acknowledgments

We acknowledge the Health Information Systems Group (SICA-CTS-553) (University of Cádiz), TIC Department of Universidad Loyola Andalucía (Sevilla), and Guadaltech Soluciones Tecnológicas (Sevilla). We specially thank the contribution of the late I Fernandez de Bobadilla to this study. Other members of the DESDE-AND Group are A Rabadán, A Mora, A García-Lupato, C Bejar Reyes, C García-Alonso, C García Pérez, C López Mata, D Jiménez-Calvo, E Motrico, FJ Delgado Leal, F Romero-Alfaro, FJ Ortiz, I González Seco, I Carmona, J Almenara, JA García-Morales, JA Galera, JD Soto-Alba, J Limón, J Álvarez, M Poole, M Martínez-Domene, MC González-Montero, MF Raposo Triano, MJ Abellán, M Rabadán, P Recio, P Sánchez Villegas, R Vaquero López, R Vera Gálvez, R López-Medel, and S Rodríguez Gómez. We also acknowledge Assistant Professor T. Chen (Canberra University, Australia) for his advice and feedback on co-design and co-creation. The DESDE-AND project has been funded by the Andalusian Regional Government: call for biomedical and health sciences research (BOJA 77), Ref. PI-0075: "DESDE AND: Adaptación, validación y pilotaje del instrumento DESDE-LTC para la clasificación estandarizada de servicios de atención sociosanitaria de la Comunidad Autónoma de Andalucía".

Authors' Contributions

All authors have participated in different parts of the project DESDE-AND: LS-C and FA-T carried out the design and coordination of the DESDE-AND project; JAS-P, JLA-A, SPP, J-LG-C, CR-L-A, and MRG-C participated in the revision of the project' materials; and CR-L-A and LS-C in the writing of the final manuscript. All authors read, contributed, and approved the final version of this text.

Conflicts of Interest

The authors are members of Psicost, a nonprofit research organization that shares the Official Registry of Computer Program Authorship of DESDE-AND together with the Government of Andalucia (Junta de Andalucía).

References

1. Salvador-Carulla L, Rosenberg S, Mendoza J, Tabatabaei-Jafari H, Pandemic-Mental Health International Network. Rapid response to crisis: Health system lessons from the active period of COVID-19. Health Policy Technol 2020 Dec;9(4):578-586 [FREE Full text] [doi: 10.1016/j.hlpt.2020.08.011] [Medline: 32874862]
2. Ouslander JG, Grabowski DC. COVID-19 in Nursing Homes: Calming the Perfect Storm. J Am Geriatr Soc 2020 Oct;68(10):2153-2162. [doi: 10.1111/jgs.16784] [Medline: 32735036]
3. Pokhilenko I, Janssen L, Evers SM, Drost R, Simon J, König HH, et al. Exploring the identification, validation, and categorization of costs and benefits of education in mental health: The PECUNIA project. Int J Technol Assess Health Care 2020 May 28;36(4):325-331. [doi: 10.1017/S0266462320000203]
4. Janssen L, Pokhilenko I, Evers S, Paulus A, Simon J, König H, et al. Exploring the identification, validation, and categorization of the cost and benefits of criminal justice in mental health: the PECUNIA project. Exploring the identification, validation, and categorization of costs and benefits of education in mental health: The PECUNIA project. Int J Technol Assess Health Care 2020;(1?8) 2020 Jul 27;36(4):418-425. [doi: 10.1017/S0266462320000471] [Medline: 32715991]
5. Pithara C, Farr M, Sullivan SA, Edwards HB, Hall W, Gadd C, et al. Implementing a Digital Tool to Support Shared Care Planning in Community-Based Mental Health Services: Qualitative Evaluation. J Med Internet Res 2020 Mar 19;22(3):e14868 [FREE Full text] [doi: 10.2196/14868] [Medline: 32191210]

6. Furst M, Bagheri N, Salvador-carulla L. An ecosystems approach to mental health services research. BJPsych International 2021 Mar 24;18(1):23-25 [FREE Full text] [doi: 10.1192/bji.2020.24]
7. Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 7. Deciding what evidence to include. Health Res Policy Syst 2006 Dec 01;4:19 [FREE Full text] [doi: 10.1186/1478-4505-4-19] [Medline: 17140445]
8. Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence. Health Res Policy Syst 2006 Dec 05;4:20 [FREE Full text] [doi: 10.1186/1478-4505-4-20] [Medline: 17147809]
9. Eurostat. Statistical Yearbook on Candidate Countries: Data 1997-2001. Luxembourg: Office for Official Publications of the European Communities; 2003. URL: https://ec.europa.eu/eurostat/web/products-statistical-books/-/KS-AG-03-001 [accessed 2021-02-22]
10. Salvador-Carulla L, Poole M, Gonzalez-Caballero JL, Romero C, Salinas JA, Lagares-Franco CM, RIRAG/PSICOST Group, DESDE Consensus Panel. Development and usefulness of an instrument for the standard description and comparison of services for disabilities (DESDE). Acta Psychiatr Scand Suppl 2006(432):19-28. [doi: 10.1111/j.1600-0447.2006.00916.x] [Medline: 17087812]
11. Johri M, Beland F, Bergman H. International experiments in integrated care for the elderly: a synthesis of the evidence. Int J Geriatr Psychiatry 2003 Mar;18(3):222-235. [doi: 10.1002/gps.819] [Medline: 12642892]
12. Salvador-Carulla L, Romero C, Weber G, Dimitrov H, Sprah L, Venner B, eDESDE-LTC Group. Classification, assessment and comparison of European LTC services Development of an integrated system. Ageing and long-term care 2011;17(2-3):27-29.
13. Johnson S, Salvador-Carulla L. Description and classification of mental health services: a European perspective. Eur. psychiatr 1998;13(7):333-341. [doi: 10.1016/S0924-9338(99)80699-3] [Medline: 19706262]
14. Salvador-Carulla L, Amaddeo F, Gutiérrez-Colosía MR, Salazzari D, Gonzalez-Caballero JL, Montagni I, et al. Developing a tool for mapping adult mental health care provision in Europe: the REMAST research protocol and its contribution to better integrated care. Int J Integr Care 2015;15:e042 [FREE Full text] [doi: 10.5334/ijic.2417] [Medline: 27118959]
15. Salvador-Carulla L, Alvarez-Galvez J, Romero C, Gutiérrez-Colosía MR, Weber G, McDaid D, et al. Evaluation of an integrated system for classification, assessment and comparison of services for long-term care in Europe: the eDESDE-LTC study. BMC Health Serv Res 2013 Jun 15;13(218):218-212 [FREE Full text] [doi: 10.1186/1472-6963-13-218] [Medline: 23768163]
16. Fleury M. Quebec mental health services networks: models and implementation. Int J Integr Care 2005;5:e07 [FREE Full text] [doi: 10.5334/ijic.127] [Medline: 16773157]
17. Madden R, Sykes C, Ustun TB. World Health Organization family of international classifications: definition, scope and purpose. Geneva, Switzerland: World Health Organization; 2007.
18. Salvador-Carulla L, Dimitrov H, Weber G, McDaid D, Venner B, Sprah L, eDESDE-LTC Group. An integrated system for the evaluation and classification of health and social services for persons with long term care needs in Europe. Barcelona: PSICOST and Catalunya Caixa; 2011.
19. Salinas-Pérez JA, Gutiérrez-Colosia MR, Romero López-Alberca C, Poole M, Rodero-Cosano ML, García-Alonso CR, et al. [Everything is on the map: Integrated Mental Health Atlases as support tools for service planning. SESPAS Report 2020]. Gac Sanit 2020 Sep;34 Suppl 1:11-19 [FREE Full text] [doi: 10.1016/j.gaceta.2020.06.015] [Medline: 32933792]
20. Furst MA, Gandré C, Romero López-Alberca C, Salvador-Carulla L. Healthcare ecosystems research in mental health: a scoping review of methods to describe the context of local care delivery. BMC Health Serv Res 2019 Mar 18;19(173):1-13 [FREE Full text] [doi: 10.1186/s12913-019-4005-5] [Medline: 30885186]
21. Chisholm D, Amir M, Fleck M, Herrman H, Lomachenkov A, Lucas R, et al. Longitudinal Investigation of Depression Outcomes (the LIDO study) in primary care in six countries: comparative assessment of local health systems and resource utilization. Int. J. Method. Psychiat. Res 2006 Jun;10(2):59-71. [doi: 10.1002/mpr.101]
22. Gutiérrez-Colosía MR, Salvador-Carulla L, Salinas-Pérez JA, García-Alonso CR, Cid J, Salazzari D, REFINEMENT Group. Standard comparison of local mental health care systems in eight European countries. Epidemiol Psychiatr Sci 2019 Apr;28(2):210-223 [FREE Full text] [doi: 10.1017/S2045796017000415] [Medline: 28918762]
23. Salvador-Carulla L, Martinez-Leal R, Poole M, Salinas-Perez JA, Tamarit J, Garcia-Ibanez J, et al. Perspectives: the mental health care gap in intellectual disabilities in Spain: impact analysis and knowledge-to-action plan. J Ment Health Policy Econ 2013 Sep;16(3):131-141. [Medline: 24327483]
24. Dahl HM, Rezvyy G, Bogdanov A, Øiesvold T. Outpatient clinics treating substance use disorders in Northwest Russia and Northern Norway: a descriptive comparative study. International Journal of Circumpolar Health 2017 Dec 15;76(1):1411733. [doi: 10.1080/22423982.2017.1411733] [Medline: 29241406]
25. Ala-Nikkola T, Pirkola S, Kaila M, Joffe G, Kontio R, Oranta O, et al. Identifying Local and Centralized Mental Health Services—The Development of a New Categorizing Variable. IJERPH 2018 May 31;15(6):1131. [doi: 10.3390/ijerph15061131] [Medline: 29857540]
26. Salvador-Carulla L. Descripción estandarizada de los servicios de discapacidad para ancianos en España. Madrid: Estudios I+D+I. IMSERSO; 2003.

27. Romero-López-Alberca C, Gutiérrez-Colosía MR, Salinas-Pérez J, Almeda N, Furst M, Johnson S, et al. Standardised description of health and social care: A systematic review of use of the ESMS/DESDE (European Service Mapping Schedule/Description and Evaluation of Services and DirectoriEs). Eur Psychiatry 2019 Sep;61:97-110. [doi: 10.1016/j.eurpsy.2019.07.003] [Medline: 31426008]
28. García-Alonso CR, Almeda N, Salinas-Pérez JA, Gutiérrez-Colosía MR, Uriarte-Uriarte JJ, Salvador-Carulla L. A decision support system for assessing management interventions in a mental health ecosystem: The case of Bizkaia (Basque Country, Spain). PLoS ONE 2019 Feb 14;14(2):e0212179. [doi: 10.1371/journal.pone.0212179] [Medline: 30763361]
29. Rock D, Cross SP. Regional planning for meaningful person-centred care in mental health: context is the signal not the noise. Epidemiol Psychiatr Sci 2020 Feb 24;29(e104):1-6. [doi: 10.1017/s2045796020000153] [Medline: 32089149]
30. Chung Y, Salvador-Carulla L, Salinas-Pérez JA, Uriarte-Uriarte JJ, Iruin-Sanz A, García-Alonso CR. Use of the self-organising map network (SOMNet) as a decision support system for regional mental health planning. Health Res Policy Sys 2018 Apr 25;16(35):1-17. [doi: 10.1186/s12961-018-0308-y] [Medline: 29695248]
31. Fernandez A, Gillespie JA, Smith-Merry J, Feng X, Astell-Burt T, Maas C, et al. Integrated mental health atlas of the Western Sydney Local Health District: gaps and recommendations. Aust. Health Review 2017;41(1):38-44. [doi: 10.1071/ah15154] [Medline: 27007640]
32. Paulk MC, Weber CV, Curtis B, Chrissis MB. The Capability Maturity Model: Guidelines for Improving the Software Process. Pittsburgh, Pennsylvania: Software Engineering Institute. Carnegie Mellon University; 1995.
33. Yeh KB, Adams ML, Marshall ES, Dasgupta D, Zhunushov A, Richards AL, et al. Applying a Capability Maturity Model (CMM) to evaluate global health security-related research programmes in under-resourced areas. Global Security: Health, Science and Policy 2017 Jan 20;2(1):1-9. [doi: 10.1080/23779497.2017.1279022]
34. Consejería para la Igualdad y Bienestar Social. Centros de Servicios Sociales en Andalucía. Sevilla (Spain): Consejería para la Igualdad y el Bienestar Social de la Junta de Andalucía; 2004:1-342.
35. Law 6/2016 of 27th December, Andalusian Social Services. URL: https://www.boe.es/eli/es-an/l/2016/12/27/9
36. Johnson S, Kuhlmann R, EPCAT Group. The European Service Mapping Schedule (ESMS): development of an instrument for the description and classification of mental health services. Acta Psychiatr Scand 2000 Oct;102(Supplement 405):14-23. [doi: 10.1111/j.0902-4441.2000.t01-1-acp28-03.x]
37. Salvador-Carulla L, Romero C, Martinez A, Haro JM, Bustillo G, Ferreira A, PSICOST Group. Assessment instruments: standardization of the European Service Mapping Schedule (ESMS) in Spain. Acta Psychiatr Scand Suppl 2000;102(405):24-32. [Medline: 11129095]
38. Saeki M. Embedding Metrics into Information Systems Development Methods: An Application of Method Engineering Technique. In: Eder J, Missikoff M, editors. Advanced Information Systems Engineering. CAiSE 2003. Lecture Notes in Computer Science. Berlin: Springer; 2003.
39. European Commission. Horizon 2020. Work Programme 2014-2015. General Annexes. 2013. URL: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/reference-documents;programCode=H2020
40. Mendoza J, Wands M, Salvador-Carulla L, Hackett M, Najlepszy LF. Evaluation of the Queensland Ed-LinQ Initiative: a school-community mental health initiative. Report for the Queensland Mental Health Commission. Brisbane: ConNetica, Qld; 2014. [doi: https://doi.org/10.13140/RG.2.2.31585.81762]
41. Lukersmith S, Millington M, Salvador-Carulla L. What is Case Management? A Scoping and Mapping Review. Int J Integr Care 2016 Oct 19;16(4):1-13. [doi: 10.5334/ijic.2477]
42. Alonso-Trujillo F, Salinas-Pérez JA, Gutiérrez-Colosía MR, González-Caballero JL, Pinzón Pulido S, Jiménez González S, et al. [Impact assessment of a multisectoral plan for the promotion of health and social wellbeing in Andalusia (Spain)]. Gac Sanit 2020;34(6):615-623 [FREE Full text] [doi: 10.1016/j.gaceta.2019.01.001] [Medline: 30827502]
43. Pan SQ, Vega M, Vella AJ, Archer BH, Parlett GR. A mini-Delphi approach: An improvement on single round techniques. Progr. Tourism Hospit. Res 1996;2(1):27-39. [doi: 10.1002/(SICI)1099-1603(199603)2:1<27::AID-PTH29>3.0.CO;2-P0.co;2-p]
44. Chung Y, Bagheri N, Salinas-Perez JA, Smurthwaite K, Walsh E, Furst M, et al. Role of visual analytics in supporting mental healthcare systems research and policy: A systematic scoping review. International Journal of Information Management 2020 Feb;50:17-27. [doi: 10.1016/j.ijinfomgt.2019.04.012]
45. Psicost Scientific Association. DESDE-AND Internet. Google My Maps. URL: https://www.google.com/maps/d/viewer?mid=10uJJvDyK2DkuY2SAW3SbpWvuOGo&ll=37.380402281588175%2C-5.965878028613365&z=11
46. Tu SW, Nyulas CI, Tudorache T, Musen MA, Martinuzzi A, van Gool C, et al. Toward a Harmonized WHO Family of International Classifications Content Model. Stud Health Technol Inform 2020 Jun 16;270:1409-1410. [doi: 10.3233/SHTI200466] [Medline: 32570683]
47. Schumacher R, Lowry SZ. NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records (No. NIST Interagency/Internal Report NISTIR-7741). United States of America: Department of Commerce; 2010.
48. Chen T, Drennan J, Andrews L, Hollebeek LD. User experience sharing. EJM 2018 May 14;52(5/6):1154-1184. [doi: 10.1108/ejm-05-2016-0298]
49. Donetto S, Pierri P, Tsianakas V, Robert G. Experience-based Co-design and Healthcare Improvement: Realizing Participatory Design in the Public Sector. The Design Journal 2015 May 07;18(2):227-248. [doi: 10.2752/175630615X14212498964312]

50. Deserti A, Rizzo F, Smallman M. Experimenting with co-design in STI policy making. Policy Design and Practice 2020 May 23;3(2):135-149. [doi: 10.1080/25741292.2020.1764692]
51. Semrau M, Alem A, Abdulmalik J, Docrat S, Evans-Lacko S, Gureje O, et al. Developing capacity-building activities for mental health system strengthening in low- and middle-income countries for service users and caregivers, service planners, and researchers. Epidemiol Psychiatr Sci 2018 Feb;27(1):11-21 [FREE Full text] [doi: 10.1017/S2045796017000452] [Medline: 28965528]
52. Fletcher-Brown J, Turnbull S, Viglia G, Chen T, Pereira V. Vulnerable consumer engagement: How corporate social media can facilitate the replenishment of depleted resources. International Journal of Research in Marketing 2020:1-12 (forthcoming). [doi: 10.1016/j.ijresmar.2020.06.002]
53. Salinas-Perez JA, Salvador-Carulla L, Saldivia S, Grandon P, Minoletti A, Romero Lopez-Alberca C. Integrated mapping of local mental health systems in Central Chile. Rev Panam Salud Publica 2018;42(e144):1-12. [doi: 10.26633/rpsp.2018.144]
54. Bagheri N, Batterham PJ, Salvador-Carulla L, Chen Y, Page A, Calear AL, et al. Development of the Australian neighborhood social fragmentation index and its association with spatial variation in depression across communities. Soc Psychiatry Psychiatr Epidemiol 2019 Apr 15;54(10):1189-1198. [doi: 10.1007/s00127-019-01712-y]
55. Pirkola S, Sund R, Sailas E, Wahlbeck K. Community mental-health services and suicide rate in Finland: a nationwide small-area analysis. Lancet 2009 Jan 10;373(9658):147-153. [doi: 10.1016/S0140-6736(08)61848-6] [Medline: 19097638]
56. Roick C, Heider D, Toumi M, Angermeyer MC. The impact of caregivers' characteristics, patients' conditions and regional differences on family burden in schizophrenia: a longitudinal analysis. Acta Psychiatr Scand 2006 Nov;114(5):363-374. [doi: 10.1111/j.1600-0447.2006.00797.x] [Medline: 17022797]
57. Marwaha S, Johnson S, Bebbington P, Stafford M, Angermeyer MC, Brugha T, et al. Rates and correlates of employment in people with schizophrenia in the UK, France and Germany. Br J Psychiatry 2007 Jul;191:30-37. [doi: 10.1192/bjp.bp.105.020982] [Medline: 17602122]
58. Salize HJ, McCabe R, Bullenkamp J, Hansson L, Lauber C, Martinez-Leal R, et al. Cost of treatment of schizophrenia in six European countries. Schizophr Res 2009 Jun;111(1-3):70-77. [doi: 10.1016/j.schres.2009.03.027] [Medline: 19401265]
59. Fernandez A, Salinas-Perez JA, Gutierrez-Colosia MR, Prat-Pubill B, Serrano-Blanco A, Molina C, et al. Use of an integrated Atlas of Mental Health Care for evidence informed policy in Catalonia (Spain). Epidemiol Psychiatr Sci 2014 Sep 16;24(6):512-524. [doi: 10.1017/s2045796014000511] [Medline: 25226091]
60. Sadeniemi M, Almeda N, Salinas-Pérez JA, Gutiérrez-Colosía MR, García-Alonso C, Ala-Nikkola T, et al. A Comparison of Mental Health Care Systems in Northern and Southern Europe: A Service Mapping Study. IJERPH 2018 May 31;15(6):1133. [doi: 10.3390/ijerph15061133]
61. van Spijker BA, Salinas-Perez JA, Mendoza J, Bell T, Bagheri N, Furst MA, et al. Service availability and capacity in rural mental health in Australia: Analysing gaps using an Integrated Mental Health Atlas. Aust N Z J Psychiatry 2019 Jun 28;53(10):1000-1012. [doi: 10.1177/0004867419857809] [Medline: 31250654]
62. Consejería de Igualdad, Políticas Sociales y Conciliación. Junta de Andalucía. Digital Map of Social Services of Andalusia. URL: https://www.serviciossocialesandalucia.es/mapa-servicios-sociales
63. Lopez Segui F, Pratdepadua Bufill C, Abdon Gimenez N, Martinez Roldan J, Garcia Cuyas F. The Prescription of Mobile Apps by Primary Care Teams: A Pilot Project in Catalonia. JMIR Mhealth Uhealth 2018 Jun 21;6(6):e10701 [FREE Full text] [doi: 10.2196/10701] [Medline: 29929948]
64. Mathieson A, Grande G, Luker K. Strategies, facilitators and barriers to implementation of evidence-based practice in community nursing: a systematic mixed-studies review and qualitative synthesis. Prim Health Care Res Dev 2019 Jan;20:e6 [FREE Full text] [doi: 10.1017/S1463423618000488] [Medline: 30068402]
65. Salvador-Carulla L, Fernandez A, Madden R, Lukersmith S, Colagiuri R, Torkfar G, et al. Framing of scientific knowledge as a new category of health care research. J Eval Clin Pract 2014 Dec;20(6):1045-1055. [doi: 10.1111/jep.12286] [Medline: 25421111]
66. Health Quality Ontario. Electronic tools for health information exchange: an evidence-based analysis. Ont Health Technol Assess Ser 2013;13(11):1-76 [FREE Full text] [Medline: 24194799]
67. Chouvarda I, Maramis C, Livitckaia K, Trajkovik V, Burmaoglu S, Belani H, ENJECT Working Group 1 Network. Connected Health Services: Framework for an Impact Assessment. J Med Internet Res 2019 Sep 03;21(9):e14005 [FREE Full text] [doi: 10.2196/14005] [Medline: 31482857]

Abbreviations

AIL: Adoption Impact Ladder
DESDE-LTC: Description and Evaluation of Services and DirectoriEs for Long-Term Care
IMM: implementation maturity model
TRL: technology readiness level

Edited by R Kukafka; submitted 10.10.20; peer-reviewed by A Serrano-Blanco, B Gee; comments to author 20.10.20; revised version received 19.11.20; accepted 17.12.20; published 15.03.21.

Please cite as:
Romero-Lopez-Alberca C, Alonso-Trujillo F, Almenara-Abellan JL, Salinas-Perez JA, Gutierrez-Colosia MR, Gonzalez-Caballero JL, Pinzon Pulido S, Salvador-Carulla L
A Semiautomated Classification System for Producing Service Directories in Social and Health Care (DESDE-AND): Maturity Assessment Study
J Med Internet Res 2021;23(3):e24930
URL: https://www.jmir.org/2021/3/e24930
doi:10.2196/24930
PMID:33720035

Original Paper

Effects of Information Architecture on the Effectiveness and User Experience of Web-Based Patient Education in Middle-Aged and Older Adults: Online Randomized Experiment

Tessa Dekkers[1,2], PhD; Marijke Melles[1], PhD; Stephan B W Vehmeijer[3], PhD, MD; Huib de Ridder[1], PhD

[1]Faculty of Industrial Design Engineering, Delft University of Technology, Delft, Netherlands

[2]Faculty of Behavioural, Management and Social sciences, University of Twente, Enschede, Netherlands

[3]Department of Orthopaedic Surgery, Reinier de Graaf Hospital, Delft, Netherlands

Corresponding Author:
Tessa Dekkers, PhD
Faculty of Behavioural, Management and Social sciences
University of Twente
Drienerlolaan 5
Enschede, 7522 NB
Netherlands
Phone: 31 534899741
Email: t.dekkers@utwente.nl

Abstract

Background: Web-based patient education is increasingly offered to improve patients' ability to learn, remember, and apply health information. Efficient organization, display, and structural design, that is, information architecture (IA), can support patients' ability to independently use web-based patient education. However, the role of IA in the context of web-based patient education has not been examined systematically.

Objective: To support intervention designers in making informed choices that enhance patients' learning, this paper describes a randomized experiment on the effects of IA on the effectiveness, use, and user experience of a patient education website and examines the theoretical mechanisms that explain these effects.

Methods: Middle-aged and older adults with self-reported hip or knee joint complaints were recruited to use and evaluate 1 of 3 patient education websites containing information on total joint replacement surgery. Each website contained the same textual content based on an existing leaflet but differed in the employed IA design (tunnel, hierarchical, or matrix design). Participants rated the websites on satisfaction, engagement, control, relevance, trust, and novelty and completed an objective knowledge test. Analyses of variance and structural equation modeling were used to examine the effects of IA and construct a theoretical model.

Results: We included 215 participants in our analysis. IA did not affect knowledge gain (P=.36) or overall satisfaction (P=.07) directly. However, tunnel (mean 3.22, SD 0.67) and matrix (mean 3.17, SD 0.69) architectures were found to provide more emotional support compared with hierarchical architectures (mean 2.86, SD 0.60; P=.002). Furthermore, increased perceptions of personal relevance in the tunnel IA (β=.18) were found to improve satisfaction (β=.17) indirectly. Increased perceptions of active control in the matrix IA (β=.11) also improved satisfaction (β=.27) indirectly. The final model of the IA effects explained 74.3% of the variance in satisfaction and 6.8% of the variance in knowledge and achieved excellent fit ($\chi^2_{17,215}$=14.7; P=.62; root mean square error of approximation=0.000; 95% CI [0.000-0.053]; comparative fit index=1.00; standardized root mean square residual=0.044).

Conclusions: IA has small but notable effects on users' experiences with web-based health education interventions. Web-based patient education designers can employ tunnel IA designs to guide users through sequentially ordered content or matrix IA to offer users more control over navigation. Both improve user satisfaction by increasing user perceptions of relevance (tunnel) and active control (matrix). Although additional research is needed, hierarchical IA designs are currently not recommended, as hierarchical content is perceived as less supportive, engaging, and relevant, which may diminish the use and, in turn, the effect of the educational intervention.

(J Med Internet Res 2021;23(3):e15846) doi:10.2196/15846

KEYWORDS

user-computer interface; total joint replacement; user-centered design; health education; mobile phone; computer-assisted instruction; patient education as topic; models, theoretical; middle aged; aged; humans; internet

Introduction

Background

Verbal and written patient education methods are often supplemented with web-based education to improve patients' ability to learn, remember, and apply health information. Such improvements are needed because patients' recall of traditional education is generally poor [1-3], which negatively affects their satisfaction with care, ability to self-manage, and emotional well-being [4,5].

There are many options to engage patients with web-based education, ranging from animations and interactive exercises to tailored health advice [6]. However, for education to be the most effective, patients must be able to use such functions independently. An efficient information architecture (IA) supports independent use [7,8], yet few studies have systematically examined IA in the context of web-based health education. To support intervention designers in making informed choices that enhance patients' learning, this paper describes a randomized experiment concerning the effect of IA on the effectiveness, use, and user experience of a patient education website and the theoretical mechanisms that explain these effects. In addition, the study explores the benefit of tailoring IA to specific user profiles.

IA

IA concerns "the structural design of a shared information environment" [9]. It describes "the way in which digital content is organized and displayed, which strongly impacts users' ability to find and use content" [10]. IA has a pervasive role in website design because it affects the user's ability to find information with no or very limited training and helps save long-term costs. Web-based environments with effective IAs are typically more scalable, easier to maintain and update, and require fewer redesigns [9]. Yet, despite the importance of IA, there is a lack of primary research that examines IA specifically in the context of web-based health education. A recent review on this subject revealed that to date, only 1 study has empirically manipulated IA in isolation from other design features [10]. This study, conducted in 2012 by Crutzen et al [11] to examine web-based hepatitis information, investigated whether providing users with the opportunity to skip pages (or not) affected website use and user perceptions of efficiency, effectiveness, and enjoyment. It was found that an architecture that provided users with less control over navigation increased both website use and knowledge gain [11]. Although this study demonstrated that IA influences web-based learning experiences, it examined only one particular IA design (the tunnel). Therefore, we argue that a more comprehensive examination of IA is required. For this purpose, we used the taxonomy of 4 archetypes of IA by Danaher et al [12,13]: tunnel, hierarchical, matrix, and hybrid architectures. Hybrid architectures mix design elements of tunnel, hierarchical, and matrix architectures. Each hybrid mix may thereby present unique advantages and disadvantages that cannot be readily understood before experimentation with the nonhybrid IA designs. Therefore, this study focuses on the three nonhybrid IA designs (ie, tunnel, hierarchical, and matrix) only. The features, advantages, and disadvantages of each design are outlined below, and additional examples of each IA design are presented in the *Methods* section of this paper.

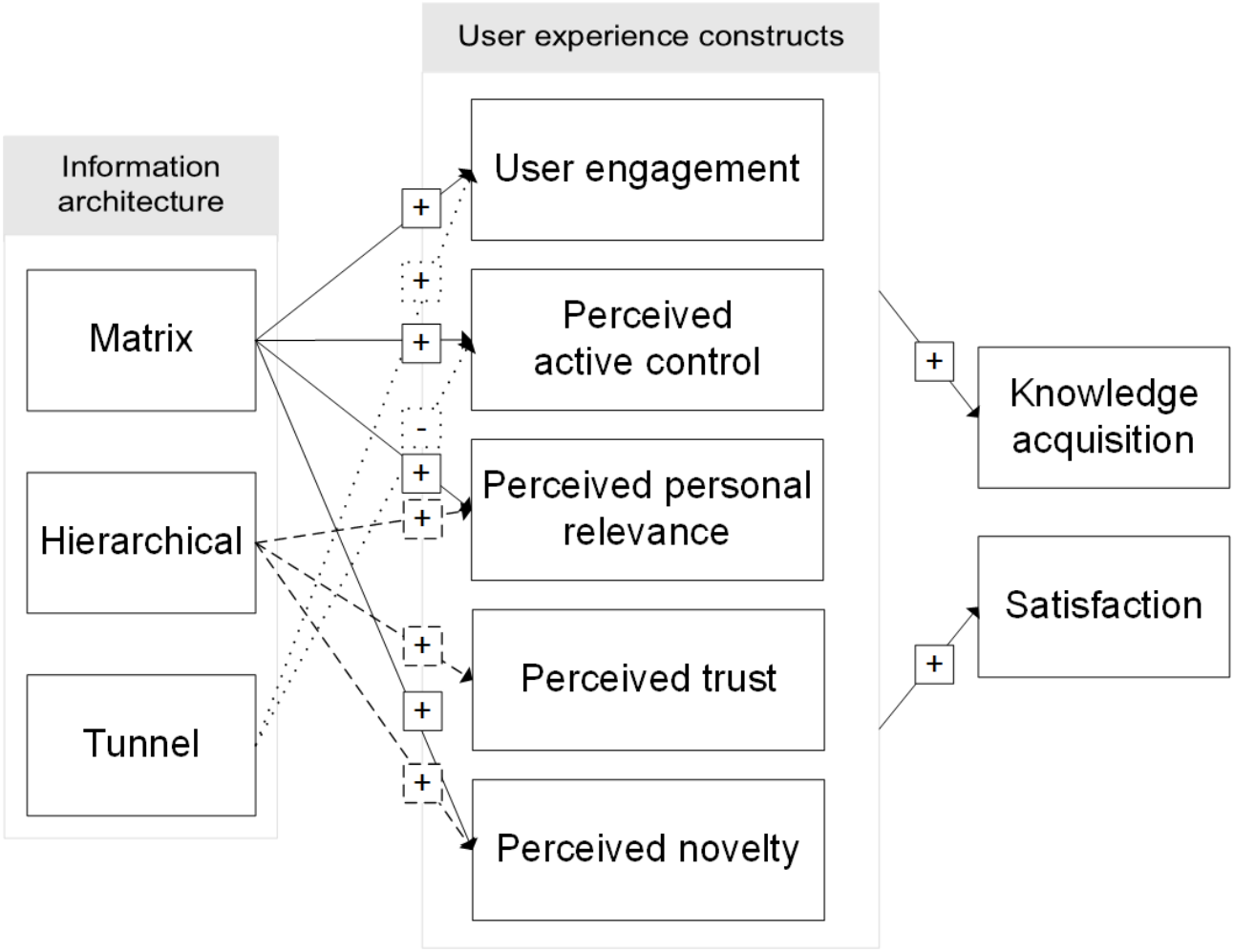
The tunnel IA design is the most common IA in health interventions: 90%-100% of interventions for chronic illness or mental health support include some form of tunneling [14]. In a typical tunnel, IA users follow a step-by-step approach to access content in a predefined, sequential order. For example, a website that only allows access to new material once users have completed previous lessons can be considered to have a tunneled design. A possible advantage of this IA is that it reduces the complexity of information. However, it also reduces the perceived control of users, which may decrease engagement and lead to nonadherence and attrition [15]. The second IA archetype is the hierarchical design. Hierarchical designs organize content hierarchically, differentiating between major and minor content. Typically, users are first provided with a general overview of the major content present on the website. For example, the official United States government website on health organizes content by major topics such as "Health Insurance," "Medications," and "Vaccines and Immunizations." After selecting the appropriate topic, users can explore nested, minor content to review in detail. Assumed advantages of this IA include increased control over content selection, familiarity, and simplicity. However, usability may be limited when users are unable to locate deeply nested content. The third IA concerns the matrix design. This IA design presents all available content on 1 home page or dashboard, thereby removing any differentiation between major and minor content or predefined sequential paths included in the hierarchical and tunnel designs, respectively. This allows users to freely navigate content in their preferred order and duration. Travel agency websites that display all available travel options first and then allow users to sort on date, price, or location are examples of matrix designs. The matrix IA design is considered engaging yet disorienting and is particularly appropriate for highly educated and experienced users looking for enrichment [15,16].

What Explains the Effects of IA?

Many scholars have condemned the *black box* approach to eHealth, which offers little understanding of the underlying mechanisms through which web-based interventions (and the tools, techniques, and strategies embedded in them) exert their effects [13,14,17]. IA design has the same issue. Although there are several assumed benefits (eg, increased usability and increased user control) of each IA design, as outlined above, there is no overarching conceptual model of IA effects. This makes it difficult to determine how IA affects the user experience of a health education website. Therefore, we examine the following 5 aspects of the user experience: user engagement, user perceptions of control, personal relevance, trustworthiness, and novelty, which may be influenced by IA design in depth.

These are depicted in the conceptual model (Figure 1). We do not hold specific expectations regarding the main effects of IA design but rather expect that each IA design may elicit a different user experience in comparison with the other IA designs, as detailed below.

Figure 1. Conceptual model of information architecture (IA). Solid arrows represent expected effects related to matrix IA design, dashed arrows represent expected effects related to hierarchical IA design, and dotted arrows represent expected effects related to tunnel IA design.



User Engagement

First, we hypothesize that IA design affects user engagement. User engagement is defined as "a quality of user experience characterized by the depth of an actor's investment when interacting with a digital system" [18,19]. It is often conceptualized as a multidimensional construct composed of cognitive, affective, and behavioral components [20], which means that engagement can both refer to a subjective experience of flow and immersion as well as the actual act of using an intervention [15]. Several recent reviews suggest that user engagement is pivotal for creating an effective and enjoyable web-based experience [15,21].

Our expectations regarding IA design as a determinant of engagement are twofold. First, tunnel IA designs (in comparison with hierarchical and matrix IA) are thought to increase behavioral engagement because the sequential, predefined setup allows researchers to persuasively guide users through the web-based process, resulting in extended use [11,14]. In a study of a web-based smoking cessation intervention, users who viewed content in a set order accessed content more often and for longer [22]. This indicates that tunnel IA design should result in higher levels of behavioral user engagement. In contrast, a more flexible matrix IA design may increase the subjective experience of engagement by providing the user more control over the interaction, as outlined below.

Perceived Active Control

As stated earlier, tunnel IA designs have been found to decrease user perceptions of control [11]. User control is a "user's ability to voluntarily participate in and instrumentally influence a communication" [23,24]. As matrix IA designs allow users to both influence the selection of content and the order in which content is consumed, this design is expected to increase perceptions of user control. Active user control is a dimension of interactivity [23,24], and interactive interventions, in turn, are associated with a more engaging experience [6,25]. Possibly, this is because users who are able to influence an intervention instrumentally consider this to be an enjoyable experience or become more emotionally invested in the intervention. It is important to note here that *perceived* interactivity and control appear to be more important than actual website interactivity [26,27]. Together, this indicates that matrix IA designs may

also improve (cognitive or affective components of) engagement through increased user perceptions of control.

Perceived Personal Relevance

Perceived personal relevance refers to the extent to which people feel that information is relevant to themselves and their situation [28-30]. People are more motivated to process personally relevant content, leading to deeper processing and greater susceptibility to any persuasive attempts the content makes [28,31,32]. Perceptions of relevance have also been linked to educational enjoyment [33]. We expect that perceived personal relevance may increase knowledge acquisition through the same motivational pathway. Hierarchical and matrix IA designs are the only designs that allow users to select content. We expect that users, to some extent, select content based on what they consider most personally relevant. Therefore, we hypothesize that hierarchical and matrix IA designs (in contrast to tunnel IA design) increase the perceived personal relevance of the health information presented and that this leads to both greater knowledge acquisition and greater satisfaction.

Perceived Trust

Perceived trust is a belief that influences whether a patient is willing to engage with health education [34]. Trust in health information is influenced by source, message, channel, and recipient [35,36] as well as structural website features [37]. A previous study on the credibility of health websites showed that the presence of a navigation menu (as is included in most hierarchical IA designs) increases perceived website credibility, as it reinforces the notion that the website is produced by a professional organization [37]. This type of heuristic evaluation of information credibility can lead to a better experience on the health website [38]. Therefore, we hypothesize that hierarchical IA design positively influences participants' trust in the health information presented and, in turn, the knowledge and satisfaction derived from the education.

Perceived Novelty

Finally, we considered perceived novelty as a potential explanatory variable. As the tunnel IA design is the norm in health interventions, other IA designs may offer more novel ways to access health information. Novelty in the context of interfaces can "act as a curiosity generating mechanism that arouses the imaginations of users and captures their interest in a site" [39]. Users pay greater attention and effort to novel media [40], subsequently leading to a greater uptake of information. Novelty has also been related to enjoyable experiences of flow and engagement [18,38]. Therefore, we expect that the less common IA designs (hierarchical and matrix) will increase user perceptions of novelty and that increased novelty will improve both user satisfaction and knowledge acquisition through increased attention to the content.

Does One IA Design Fit All?

A final consideration in examining the effects of IA is the role of individual preferences and capabilities. Many recommendations regarding IA design take user characteristics into account. For example, Lynch and Horton [16] describe matrix IA designs (which they refer to as *webs*) as more suitable for highly educated users with a high level of prior knowledge about the content. It has also been suggested that perceived control over website navigation may be more important to some users than to others [11]. However, the influence of individual differences on the effectiveness and experience of different IA designs has not been empirically tested.

This study used a previously defined set of user profiles of patients [41] who had undergone total joint replacement (TJR) surgery to explore the potential benefit of tailored IA design (Table 1). Each profile represents 1 of 3 ways through which communicative preferences and capabilities may manifest in patients. So-called *managing* patients prefer open, participative communication, particularly regarding personal circumstances, and have high capabilities and self-efficacy for understanding and applying health information. In comparison, *optimistic* patients have similar capabilities but find patient-provider communication of lesser importance and only have a slight preference for an open communicative style. Finally, *modest* patients value both open information and emotional support but have limited self-efficacy and skills in health communication. With these profiles and the recommendations for each IA design in mind, we hypothesize that users with higher preferences for open communication (ie, managing patients) will prefer IA designs that offer more control (ie, matrix), optimistic patients will not prefer any IA design in particular, and modest patients will prefer more supportive IA designs that guide them through the educational content step by step (ie, tunnel).

Table 1. Description of communicative preferences and capabilities of three total joint replacement patient profiles[a].

Managing profile	Optimistic profile	Modest profile
High preference for open communication	Moderate preference for open communication	Moderate preference for open communication
High preference for emotionally supportive communication	Low preference for emotionally supportive communication	Moderate preference for emotionally supportive communication
High critical communication capabilities	Moderate critical communication capabilities	Low critical communication skills
High personal communication capabilities	Moderate personal communication capabilities	Low personal communication skills
High self-efficacy for health information	High self-efficacy for health information	Low self-efficacy for health information

[a]Patient profiles are based on Groeneveld et al [41].

Study Objectives

The aims of this study are threefold: (1) to test the effects of IA in the context of a TJR surgery patient education website on knowledge acquisition and satisfaction with web-based education; (2) to test possible working mechanisms of IAs, including user engagement, perceived user control, perceived personal relevance, perceived trust, and perceived novelty; and (3) to explore the potential of tailored IAs.

Methods

Design

In July 2018, we conducted a between-subjects experiment comparing the knowledge and satisfaction gained from a patient education website with three different IA designs. Ethics approval for this study was obtained from the Human Research Ethics Committee Delft University of Technology. Participants provided written consent and signed a data processing agreement formulated in concordance with the General Data Protection Regulation.

Participants and Procedure

Participants were recruited using a Dutch web-based consumer research service (respondenten.nl B.V.). Middle-aged to older adults (40-80 years) with self-reported chronic hip or knee joint complaints (including arthrosis, wear and tear, chronic inflammation, birth deficits, or unknown causes) were eligible for participation. To detect a small-to-medium effect (f^2=0.15-0.25) on satisfaction and knowledge using an α of .05 and a power of 0.80, a sample size between 159 and 432 participants was needed [42,43]. We aimed to recruit at least 100 participants per condition for a total sample of 300 participants. In total, we were able to enroll 235 participants, of which the data of 215 participants were included in the analysis (see the *Results* section). Participants received monetary reimbursement (15 euro [US $18.2]) for their participation.

The complete experiment was conducted on the web via survey software (Qualtrics). Each eligible participant was provided a hyperlink to the survey. After providing consent, participants filled out questionnaires regarding their communication preferences and skills, health, anxiety, and coping behavior, which were used to determine the patient profile [41]. Participants also stated the extent to which they already felt knowledgeable about TJR surgery (part A). In part B, participants were randomly assigned to 1 of 3 experimental conditions using Qualtrics' built-in randomizer. The allocation sequence and assignments were concealed from all participants, the researchers, and the consultant hired for participant recruitment until all data were collected. Participants were initially asked to focus on either the website's design or its content. After reviewing the website's design, participants reported satisfaction and user perceptions. They were then asked to view the website for a second time while focusing on content. Then, they completed a knowledge test designed for the purpose of this study. The order of focus (design vs content) was counter-balanced. Finally, participants shared their sociodemographic information and received a code for reimbursement (part C). Eligible participants who had not started or completed the survey after 3 weeks were reminded via email once.

Materials

Design Process

The three websites were designed between March and June 2018 by a design agency (Panton B.V.) specializing in the design of products, services, and processes for health care under the supervision of the first author. The lead designer provided literature on IA [12] and was given access to patient profile role descriptions and anonymized data about patients' communication preferences and capabilities collected in an earlier study (T Dekkers, PhD, unpublished data, February 2017). In June, prototypes of the websites were pilot tested. To discuss progress and ensure accuracy and quality of health information shared on the patient education websites, the design team met with the first author 10 times throughout the design process. At 2 points in the design process (after first conceptualization and after the pilot tests), the design team also met with the full research team, including an orthopedic surgeon.

Pilot Usability Study

Prototypes of the three websites were pilot tested with 7 patients (age range 46-77 years) scheduled for TJR surgery and 7 informal caregivers (age range 42-76 years) in June 2018. The pilot test focused specifically on usability of the websites rather than effectiveness in terms of knowledge acquisition. Interested patients present at the clinic for scheduled group-based patient education were shown the prototypes after they provided written consent. They first freely explored the websites while mentioning aloud any (positive or negative) aspects that stood out. Then, they were asked to find information about the first checkup after surgery. This assignment was used to identify usability issues and software bugs [44]. Finally, patients were asked to report engagement using the User Engagement Scale-Short Form (UES-SF, see *Measurements* section). Throughout the pilot test, the cursor of the participants was tracked using screen capture software (CamStudio Recorder v2.7, Rendersoft Development). Screen captures were used both to identify unclear navigational cues and to get an initial impression of whether the users navigated through the IAs as intended (eg, whether patients explored more pages in the matrix design, made use of the table of contents in the hierarchical design, and moved step by step using the next and prior buttons in the tunnel design). The input of patients and caregivers was shared with the lead designer and implemented in the following iteration of the design. This led to significant improvements in usability, including less scrollable text, more prominently displayed contact information, vivid color accents, and larger buttons.

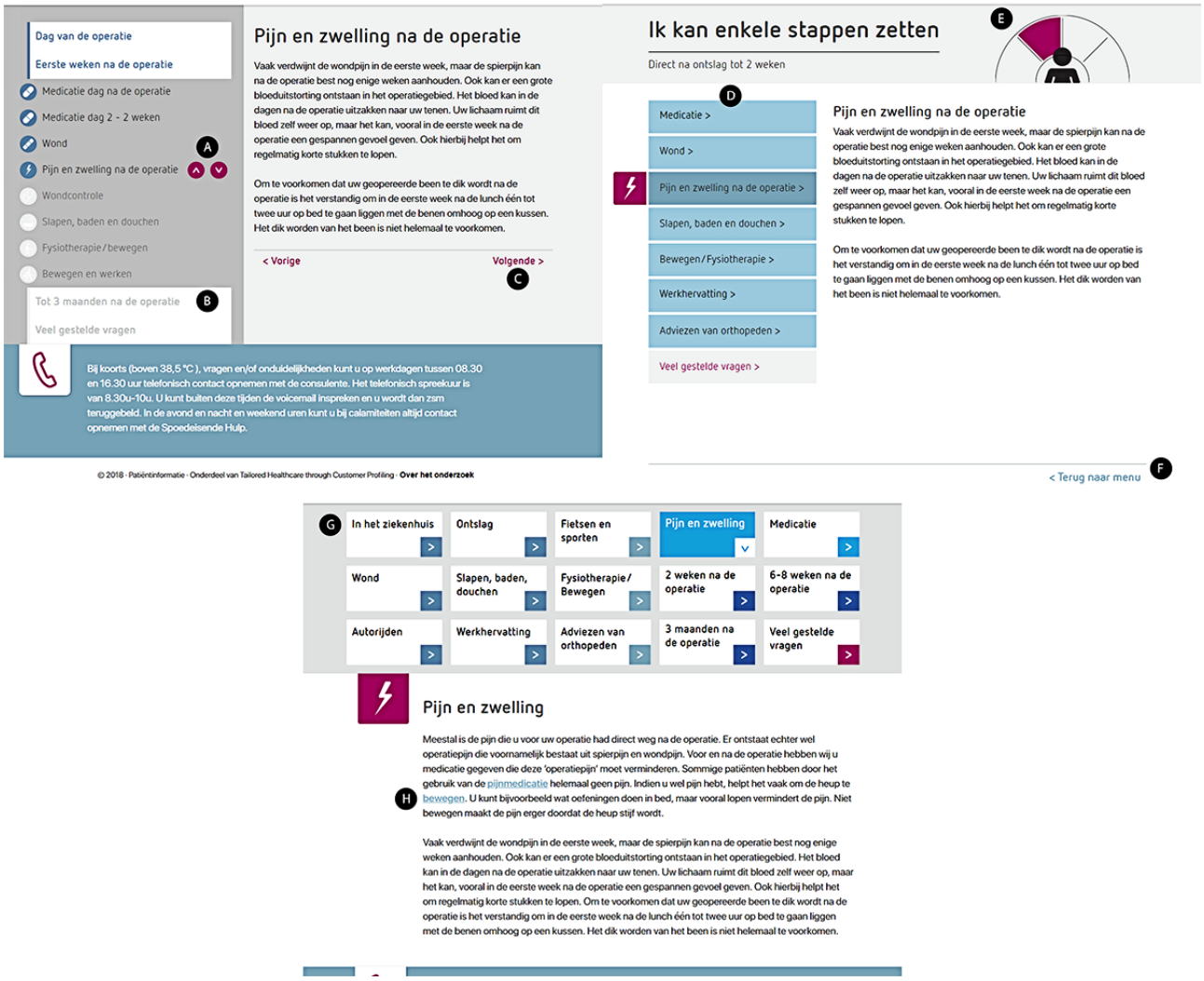
Websites

All websites contained the same textual content based on an existing patient education leaflet titled *Instructions after an outpatient Total Hip Prosthesis (THP; Instructies na een Totale Heup Prothese [THP] in dagbehandeling)* used by the local hospital (Reinier de Graaf Gasthuis, the Netherlands). The leaflet addressed practical concerns before and after outpatient THP surgery, including preparation for surgery, pain, medication,

and physiotherapy. All graphic design elements (including photos, fonts, and color) were equivalent across websites.

The tunnel IA website design had a chronological sequential ordering of topics presented as a timeline, starting with *the day of the operation* and ending with the *3-month follow-up* and *frequently asked questions*. Navigation was limited to *next* and *previous* buttons placed below the text and in the timeline. Topics that were not yet accessible to the user were grayed out (Figure 2). The hierarchical IA website design presented participants with a choice menu in which they selected the phase of their *patient journey* (eg, in the hospital and able to walk a few steps). After selecting an option, users were presented with topics grouped in a table-of-content menu. Participants could further investigate their chosen topic using the menu and could return to the home page using the buttons or navigation path (ie, *bread crumb trail*). The matrix IA website design showed all topics in tiles on the home page and provided no suggested reading order. By clicking on the topic tiles or hyperlinks in the body of text, participants could switch between topics. Offline copies of the experimental websites are available on request by contacting the first author.

Figure 2. Annotated screenshots of tunnel, hierarchical, and matrix information architecture (IA) design of a Dutch patient education website to prepare patients for total joint replacement surgery. Tunnel IA: (A) next/previous buttons, (B) grayed-out text (not yet accessible), and (C) next/previous buttons. Hierarchical IA: (D) table of contents, (E) major grouping by recovery phase, and (F) return to main menu. Matrix IA: (G) topic matrix and (H) hyperlink. All screenshots depict the same content about pain and swelling (pijn en zwelling).



Measurements

The primary outcomes of interest are knowledge acquisition and website satisfaction. Satisfaction with web-based education captures both the attitude of patients toward website functioning (eg, satisfaction with comprehensibility and with emotional support derived from the website) as well as their affective attitude (eg, satisfaction with website attractiveness) [45,46]. The secondary outcomes used to test the conceptual model include user perceptions of engagement, control, personal relevance, trust, and novelty. We also measured use by capturing the total time spent on the website in minutes. Finally, we collected short qualitative feedback forms on the perceived advantages and disadvantages of the website.

Knowledge Acquisition and Satisfaction With Website

A total of 5 multiple-choice (MC) questions and 3 open questions about (self-)care after TJR surgery were used to assess knowledge acquisition. The questions were based on the information provided on the websites and included, for example: *after the surgery, it is important to strengthen the muscles surrounding the hip joint. Which ways to do so are*

recommended by orthopedic surgeons? Each question included the following answer options: *not been discussed*, *discussed, but I cannot remember the details*, a correct answer, and an incorrect answer (distractor) [47]. For each correct MC answer, participants scored 1 point, and for each open question, an answer sheet was developed that assigned points from 0 (incorrect), 1 (partly correct), to 2 (fully correct). All points were summed and converted to reflect the percentage of correct answers (0%-100% correct).

Satisfaction with patient education was measured using the Website Satisfaction Scale [45,46] comprising three subscales: satisfaction with the (1) attractiveness, (2) comprehensibility of the information, and (3) emotional support received from the website. All items consisted of statements to which participants' agreement was measured on a 5-point Likert scale (1=*totally disagree* and 5=*totally agree*). Statements included *the website looks nice*, *the website is understandable*, and *the website give ease of mind*. Both the overall index score of satisfaction and the separate subscales achieved excellent reliability (α=.82-.98).

User Perceptions of Engagement, Active Control, Personal Relevance, Trust, and Novelty

We included 5 constructs to explore the theoretical mechanisms through which (tailored) IAs may influence knowledge acquisition and satisfaction. The first is user engagement, as measured through the UES-SF [19]. We obtained permission to translate this validated questionnaire to Dutch according to the guidelines for cross-cultural adaptation of self-reported instruments [48,49] (personal communication by HL O'Brien, May 18, 2018). The instrument contains 12 questions, which form 1 index score (α=.88), and 4 subscales: focused attention (*I was absorbed in this experience*, α=.75), aesthetic appeal (*the website was attractive*, α=.87), reward (*using the website was worthwhile*, α=.71), and perceived usability (*I felt frustrated while using the website*, α=.79; Multimedia Appendix 1 [18,19,50]). The other user perceptions of interest included perceived active control (*during the website visit, I could freely decide what I wanted to see*, 4 items, α=.96) [27], personal relevance (*the website was relevant to my situation*, 2 items, α=.83) [51], trust (*the website is sincere and honest*, 3 items, α=.97) [34], and novelty (*the website incited my curiosity*, 3 items, α=.90) [50]. All questions were answered on a 5-point Likert scale (1=*strongly disagree* and 5=*strongly agree*).

Statistical Methods

We conducted chi-square (χ^2) and analyses of variance (ANOVA) tests to check whether background characteristics were evenly distributed over experimental conditions. To test the main effect of IA, 2 ANOVA tests were conducted with satisfaction and knowledge gain as dependent variables. Follow-up pairwise *t* tests were performed to explore differences between the IA designs, and these were all corrected using the Bonferroni correction. Finally, ANOVA tests were performed with the secondary outcomes (user perceptions) as dependent variables, and the concept of tailored IAs was explored in a two-way ANOVA with condition and profile as the independent variables.

To construct a conceptual model of how IA influences satisfaction and knowledge acquisition, we used structural equation modeling. User perceptions of engagement, personal relevance, active control, trust, and novelty (hereafter, mediating variables) were regressed on IA. Satisfaction and knowledge were regressed on IA and the mediating variables. To improve the parsimony and fit of the model, we removed nonsignificant paths. As our hypotheses suggest that IA design may influence perceived control and subsequently user engagement, and ultimately satisfaction and knowledge, we also constructed a separate serial mediation model for this hypothesis specifically. Model chi-square (χ^2), comparative fit index (CFI), standardized root mean square residual (SRMR), and root mean square error of approximation (RMSEA) were used to determine model fit. A model was considered to have a good fit when χ^2 divided by degrees of freedom ≤3 with $P<.05$, CFI≥0.95, SRMR≤0.09, and RMSEA≤0.07 [52,53]. All analyses were conducted using R version 3.5.1 [54] with α=.05.

Results

Participant Characteristics

We enrolled 235 participants, of which, 215 participants were included in the analysis (Figure 3). A total of 20 participants completed the survey on a mobile device, despite instructions to view the survey and the websites on a laptop or personal computer. As the layout and, thus, the information architecture of the websites may appear distorted on mobile devices, these participants were excluded from analysis. There were no significant differences between the excluded participants compared with the included participants with respect to background characteristics, except for device use ($P<.001$). Excluded participants used the personal computer less (47% vs 9% nonuse) and tablet devices more (89% vs 41% use). No significant associations were found between background characteristics and experimental conditions, indicating that participants were evenly distributed over all three conditions. All participant characteristics are reported in Table 2. On average, participants were 57 years old (SD 7.7), female (155/215, 72.1%), attained lower secondary education (95/215, 44.2%), and were employed or self-employed (118/215, 54.9%). They used the internet daily (mean 3.2 hours, SD 2.1) mainly on personal computers or laptops (91%) and mobile phones (82%). Participants rated their overall health significantly lower (69 out of 100) than the Dutch average of 81.5 for people aged 50-59 years [55,56] and experienced considerable movement-evoked joint pain (mean 4.9, SD 2.3).

Figure 3. Participant recruitment and follow-up diagram. IA: information architecture.

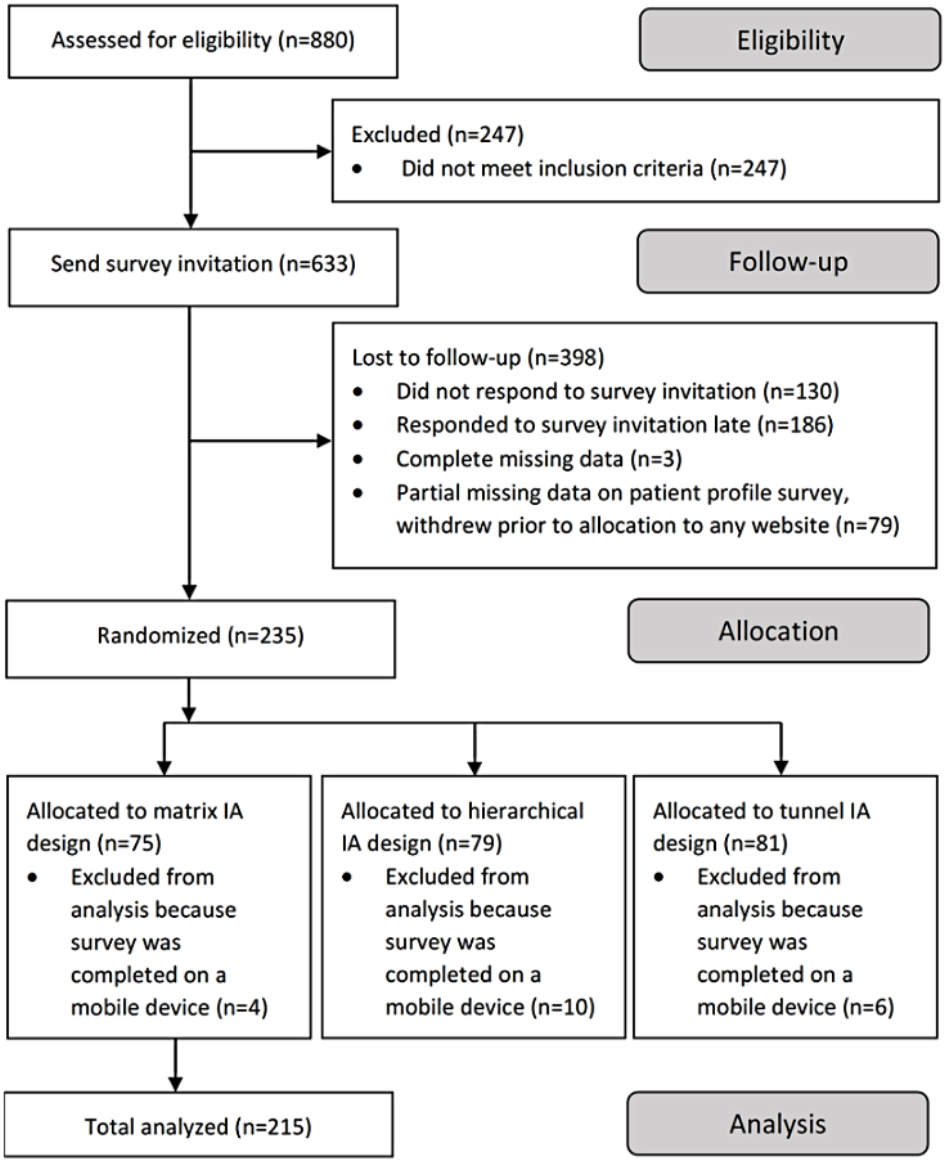


Table 2. Participant characteristics (N=215).

Variable	Value
Age (years), mean (SD)[a]	57.18 (7.70)
Sex, n (%)	
Female	155 (72.1)
Male	60 (27.9)
Education, n (%)	
Primary education	3 (1.4)
Lower secondary education	95 (44.2)
Higher secondary education	36 (16.7)
Tertiary education	81 (37.7)
Occupation, n (%)	
Employed	83 (38.6)
Self-employed	35 (16.3)
Retired	37 (17.2)
Beneficiary	29 (13.5)
Other or none	31 (14.4)
Relationship status, n (%)	
Married or long-term relationship	132 (61.4)
Divorced	41 (19.1)
Never married	35 (16.3)
Widowed	5 (2.3)
Other	2 (0.9)
Social support[b], n (%)	
Partner	124 (57.7)
Friend	75 (34.9)
Child	52 (24.2)
Neighbor	36 (16.7)
Family member	34 (15.8)
Colleague	7 (3.3)
Group (church or sports)	4 (1.9)
Other	2 (0.9)
No support	25 (11.6)
Internet use in hours per day, mean (SD)[c]	3.17 (2.14)
Device use[b], n (%)[a]	
Personal computer or laptop	194 (90.7)
Phone	175 (81.8)
Tablet	88 (41.1)
Self-reported previous knowledge of hip replacement surgery, mean (SD)[d]	1.85 (0.92)
Patient profile, n (%)	
Optimistic	90 (41.9)
Modest	72 (33.5)
Managing	53 (24.7)

[a]Data were missing for 1 participant.

[b]Participants could select multiple answers.

[c]Data were missing for 10 participants.

[d]Data were missing for 2 participants.

Effects of IA on Knowledge Acquisition and Satisfaction

All three websites received predominantly positive feedback via the open qualitative feedback forms; participants appreciated that they were *clear and organized*. Multimedia Appendix 2 summarizes the qualitative feedback on advantages and disadvantages for each IA. Table 3 and Figure 4 report the overall effects of IA. IA did not directly affect knowledge acquisition ($F_{2,212}$=1.023; P=.36; η_p^2=0.010) or overall satisfaction ($F_{2,212}$=2.702; P=.07; η^2=0.025). IA did have a significant effect on satisfaction with emotional support ($F_{2,212}$=6.376; P=.002; η^2=0.057). Post hoc analyses indicated that participants were significantly less satisfied with the hierarchical IA design (mean 2.86, SD 0.60) compared with the matrix (mean 3.17, SD 0.69) and tunnel (mean 3.22, SD 0.67) architectures. The hierarchical design was perceived as the least favorable in general: users devoted less focused attention (mean difference to tunnel −0.31; P=.03), saw the design as less novel (mean difference to tunnel −0.33; P=.02 and mean difference to matrix −0.36; P=.01) and less personally relevant (mean difference to tunnel −0.44; P=.006), and found that it provided the least active control (mean difference to matrix −0.32; P=.02).

Table 3. Knowledge acquisition, satisfaction, and user perceptions of patient education website by information architecture.

Outcome, mean (SD)	Tunnel IA[a] (n=75)	Matrix IA (n=71)	Hierarchical IA (n=69)	*P* value	(η^2)[b]
Website satisfaction	3.69 (0.52)	3.65 (0.52)	3.50 (0.48)	.07	N/A[c]
Attractiveness	3.73 (0.61)	3.68 (0.65)	3.61 (0.61)	.50	N/A
Comprehension	4.24 (0.56)	4.21 (0.59)	4.17 (0.71)	.79	N/A
Emotional support	3.22 (0.67)	3.17 (0.69)	2.86 (0.60)	.002[d]	.057
Knowledge acquisition	51.64 (19.55)	48.02 (19.75)	47.3 (19.63)	.36	N/A
User engagement	3.71 (0.55)	3.65 (0.55)	3.48 (0.57)	.047[e]	.028
Focused attention	3.16 (0.75)	3.00 (0.70)	2.85 (0.79)	.04[f]	.030
Esthetic appeal	3.76 (0.68)	3.75 (0.68)	3.52 (0.76)	.08	N/A
Reward	3.81 (0.62)	3.78 (0.57)	3.58 (0.68)	.06	N/A
Perceived usability	4.08 (0.68)	4.05 (0.78)	3.98 (0.78)	.67	N/A
Perceived active control	3.84 (0.67)	3.95 (0.65)	3.63 (0.74)	.02[g]	.035
Perceived personal relevance	3.08 (0.86)	2.73 (0.83)	2.64 (0.86)	.005[h]	.050
Perceived trustworthiness	3.94 (0.56)	3.92 (0.57)	3.78 (0.59)	.21	N/A
Perceived novelty	3.43 (0.75)	3.46 (0.73)	3.10 (0.76)	.007[i]	.046
Time spent in minutes:seconds	5:53 (4:24)	5:18 (4:15)	4:59 (4:09)	.44	N/A

[a]IA: information architecture.

[b]Effect size is only provided for significant differences.

[c]N/A: not applicable.

[d]Hierarchical IA was significantly different from both tunnel IA (P=.02) and matrix IA (P=.02).

[e]Hierarchical IA was significantly different from tunnel IA (P=.05).

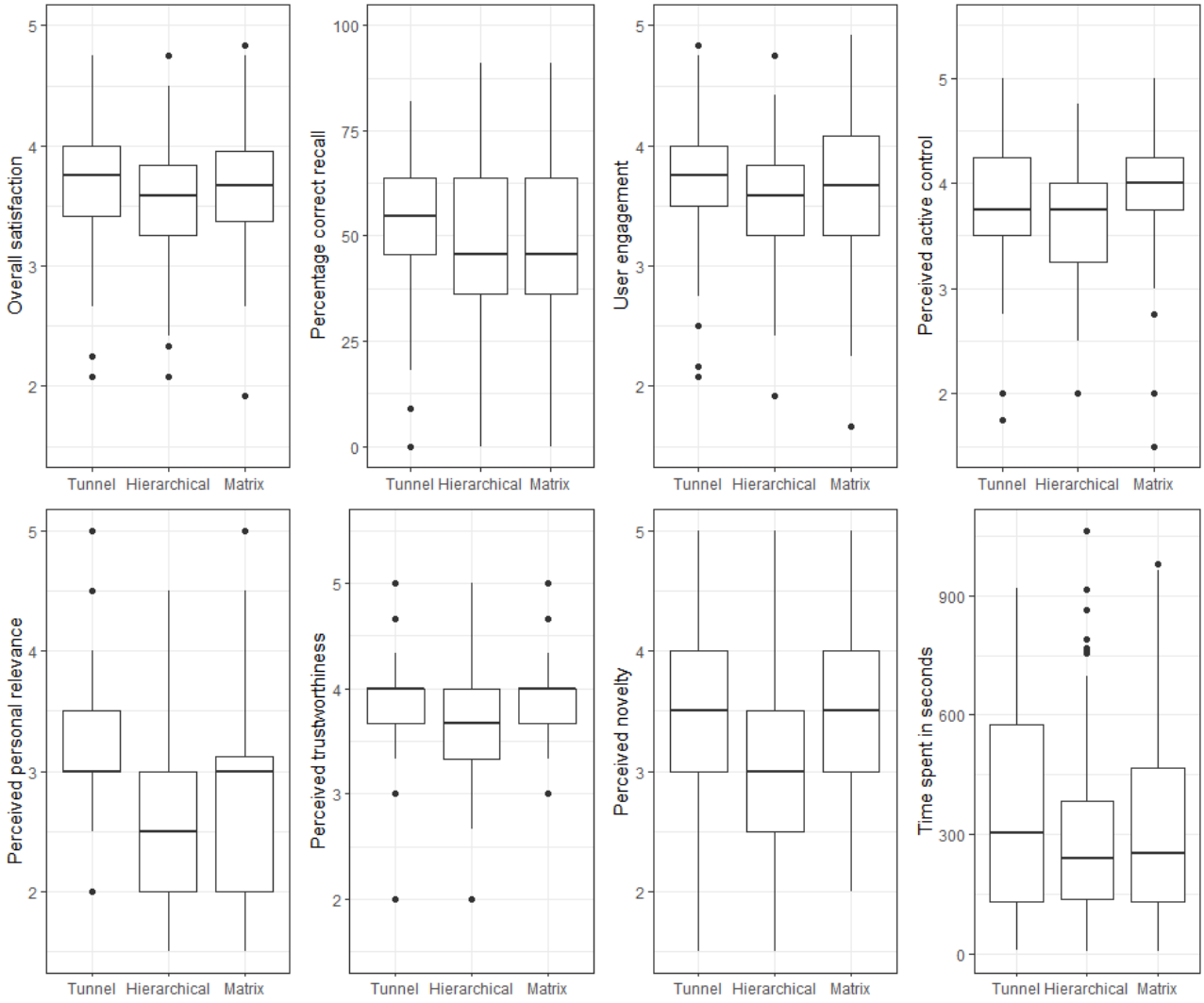
[f]Hierarchical IA was significantly different from tunnel IA (P=.03).

[g]Hierarchical IA was significantly different from matrix IA (P=.02).

[h]Tunnel IA was significantly different from both hierarchical IA (P=.006) and matrix IA (P=.04).

[i]Hierarchical IA was significantly different from both tunnel IA (P=.03) and matrix IA (P=.01).

Figure 4. Main effects of information architecture.



Model of IA Effects

The ANOVA tests demonstrated that the tunnel and matrix designs performed significantly better than the hierarchical IA design. To explain why tunnel and matrix IAs perform better compared with hierarchical IAs, we selected the hierarchical IA as the reference category in the mediation model.

The first mediation model (Model 1) specified that the effect of IA on knowledge and satisfaction would be mediated by user perceptions of engagement, active control, personal relevance, trust, and novelty. Specification of complete mediation results in a fully saturated regression model with zero degrees of freedom, as the number of observations is equal to the number of parameters [57,58]. Therefore, the first model was interpreted based on the regression paths instead of the fit indices (Table 4). All pathways (of which the exact *P* values are provided in Table 4) with *P*<.10 were considered in a second model (Model 2). For Model 3 and Model 4, we continued eliminating pathways with a more stringent cut-off of *P*<.05.

Overall, models 2 to 4 all achieved similarly good fit (Table 5). Model 4 (Figure 5) was selected as the final model, as it was the most parsimonious (expressed by highest degrees of freedom [59]). This model explained 74.3% of the variance in satisfaction and 6.8% of the variance in knowledge and achieved excellent fit ($\chi^2_{17,215}$=14.7; *P*=.62; RMSEA=0.000; CI 0.000-0.053; CFI=1.00; SRMR=0.044).

Table 4. Pathways included in mediation models 1, 2, 3, and 4.

Outcome and predictor or mediator	Path estimate (Model 1)	*P* value (Model 1)	Model 2	Model 3	Model 4
User engagement					
Tunnel IA[a]	0.190	.02	✓[b]	✓	—[c]
Matrix IA	0.139	.08	✓	—	—
Perceived active control					
Tunnel IA	0.142	.07	✓	—	—
Matrix IA	0.215	.006	✓	✓	✓
Perceived personal relevance					
Tunnel IA	0.243	.002	✓	✓	✓
Matrix IA	0.048	.54	—	—	—
Trust					
Tunnel IA	0.133	.09	✓	—	—
Matrix IA	0.109	.17	—	—	—
Perceived novelty					
Tunnel IA	0.208	.007	✓	—	—
Matrix IA	0.225	.004	✓	✓	—
Knowledge					
User engagement	0.226	.045	✓	✓	✓
Perceived active control	0.006	.96	—	—	—
Perceived personal relevance	0.089	.22	—	—	—
Trust	−0.007	.93	—	—	—
Perceived novelty	−0.006	.95	—	—	—
Satisfaction					
User engagement	0.382	<.001	✓	✓	✓
Perceived active control	0.273	<.001	✓	✓	✓
Perceived personal relevance	0.169	<.001	✓	✓	✓
Trust	0.227	<.001	✓	✓	✓
Perceived novelty	0.026	.60	—	—	—
Knowledge					
Tunnel IA design	0.042	.59	—	—	—
Matrix IA design	−0.018	.82	—	—	—
Satisfaction					
Tunnel IA design	−0.011	.80	—	—	—
Matrix IA design	−0.017	.68	—	—	—
Knowledge					
User engagement×matrix IA	0.031	.19	—	—	—
Perceived novelty×matrix IA	−0.001	.95	—	—	—
Trust×matrix IA	−0.001	.93	—	—	—
Perceived personal relevance×matrix IA	0.004	.58	—	—	—
Perceived active control×matrix IA	0.001	.96	—	—	—
User engagement×tunnel IA	0.043	.12	—	—	—
Perceived novelty×tunnel IA	−0.001	.95	—	—	—
Trust×tunnel IA	−0.001	.93	—	—	—

Outcome and predictor or mediator	Path estimate (Model 1)	*P* value (Model 1)	Model 2	Model 3	Model 4
Perceived personal relevance×tunnel IA	0.022	.25	—	—	—
Perceived active control×tunnel IA	0.001	.96	—	—	—
Satisfaction					
User engagement×tunnel IA	0.073	.02	✓	✓	—
Perceived active control×tunnel IA	0.039	.09	✓	—	—
Perceived personal relevance×tunnel IA	0.041	.01	✓	✓	✓
Trust×tunnel IA	0.030	.11	—	—	—
Perceived novelty×tunnel IA	0.005	.61	—	—	—
User engagement×matrix IA	0.053	.09	✓	—	—
Perceived active control×matrix IA	0.059	.02	✓	✓	✓
Perceived personal relevance×matrix IA	0.008	.54	—	—	—
Trust×matrix IA	0.025	.18	—	—	—
Perceived novelty×matrix IA	0.006	.61	—	—	—

[a]IA: information architecture.

[b]Pathways indicated with a check mark were included in the model formulation.

[c]Pathways indicated with an em dash were excluded in the model formulation.

Table 5. Fit statistics of mediation models 2, 3, and 4.

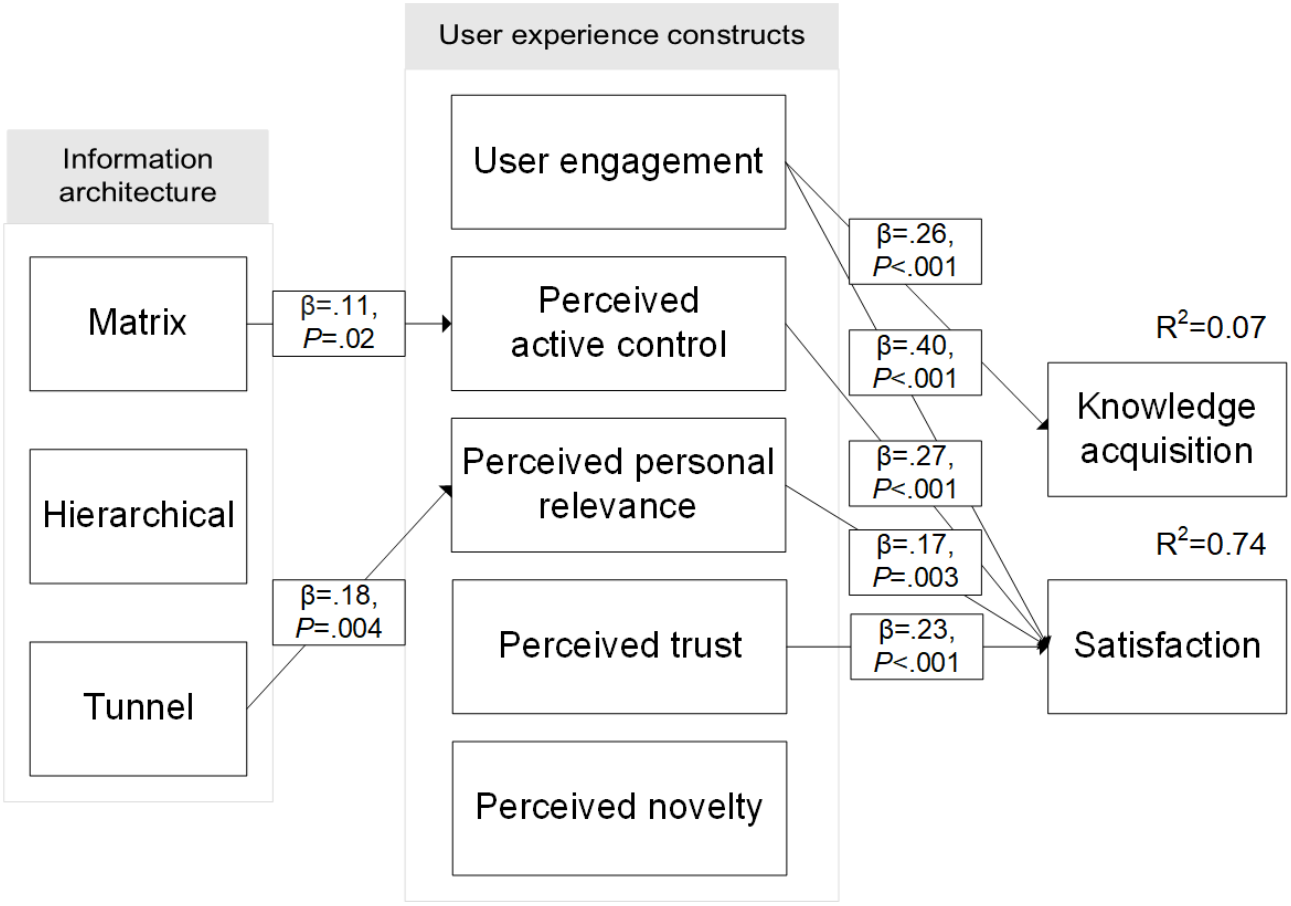
Model	Chi-square (df)	*P* value	χ^2 divided by df	CFI[a]	SRMR[b]	RMSEA[c]	95% CI
Model 2	4.7 (9)	.86	0.522	1	0.027	0.000	0.000-0.041
Model 3	10.8 (13)	.63	0.833	1	0.042	0.000	0.000-0.057
Model 4	14.7 (17)	.62	0.864	1	0.044	0.000	0.000-0.053

[a]CFI: comparative fit index.

[b]SRMR: standardized root mean square residual.

[c]RMSEA: root mean square error of approximation.

Figure 5. Structural equation model of the effects of information architecture.

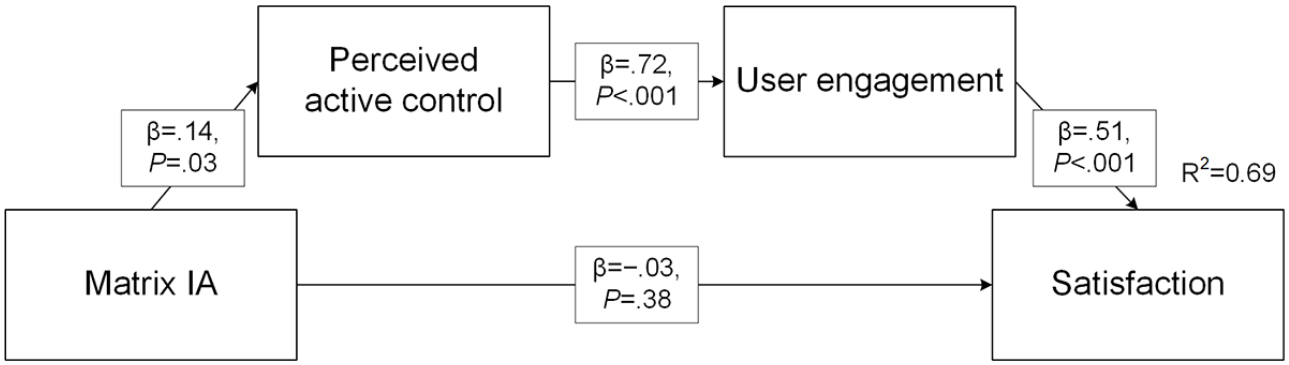


The model explains the effect of IA as follows: compared with hierarchical IAs, health information presented in a tunnel IA is perceived as more personally relevant (β=.18). This subsequently increases user satisfaction (β=.17). Matrix IAs, in comparison with hierarchical IAs, significantly increase the active control users perceive to have over the health information (β=.11), which also increases satisfaction (β=.27). Furthermore, the model shows that next to user perceptions of personal relevance and active control, user engagement and perceived trust in the health information affect users' satisfaction with a patient education website. Although we hypothesized that perceived novelty would also be affected by IA and affect satisfaction and knowledge in turn, this was not the case. Finally, we already established that IA design did not directly affect knowledge acquisition. The model demonstrated that IA also did not indirectly influence knowledge, as none of the tested mediation pathways were significant. Knowledge acquisition was influenced by user engagement (β=.26), but user engagement itself was unaffected by IA.

Serial Mediation by Perceived Control and User Engagement

The serial mediation model, including perceived control and user engagement, confirmed that IA design did not significantly predict satisfaction (P=.07) or knowledge (P=.36). However, an indirect-only serial mediation by perceived control and user engagement on satisfaction emerged for matrix IA designs (β[indirect]=.052; z=2.053; P=.04) and hierarchical designs (β[indirect]=−.063; z=−2.545; P=.01), where matrix IA increased active control and subsequently user engagement and satisfaction, whereas hierarchical design decreased active control (Figure 6). Serial mediation was not present for tunnel IA (P=.65) or for knowledge (P_{matrix}=.10, $P_{hierarchical}$=.06, P_{tunnel}=.65).

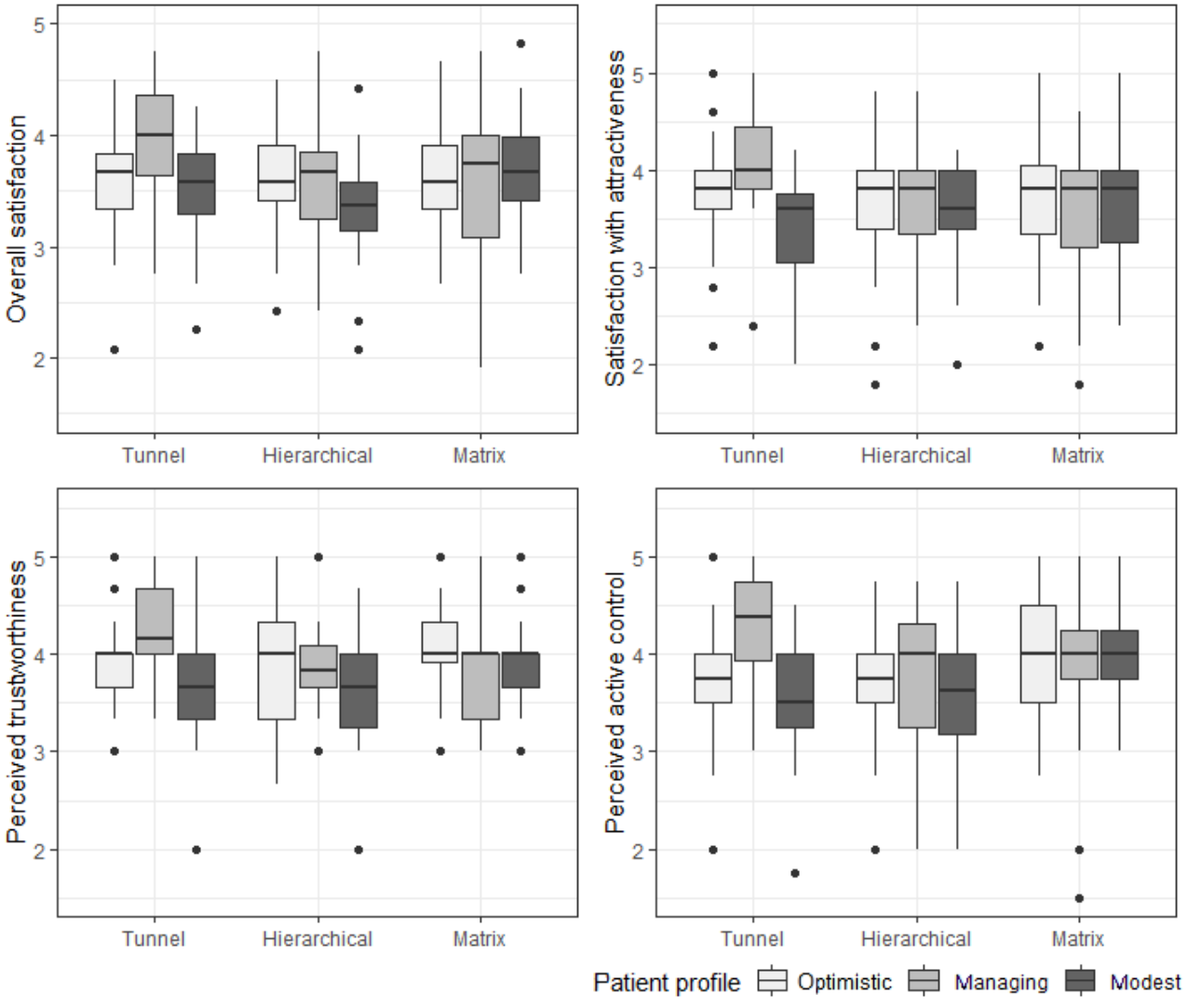
Figure 6. Serial mediation model of matrix information architecture effects on satisfaction via active control and engagement. IA: information architecture.



Tailored IAs: Interactions With Patient Profile

Interaction effects between IA and patient profile indicated that some IA designs were preferred more by users with specific profiles ($F_{4,206}$=2.646; P=.04; η_p^2=0.049). In the post hoc analyses, a consistent difference was demonstrated between participants of the managing profile and modest profile using a tunnel IA design (Figure 7). Managing participants were significantly more satisfied with the tunnel design (mean difference to modest 0.489; P=.04), perceived it as more attractive (mean difference to modest 0.673; P=.01) and trustworthy (mean difference to modest 0.630; P=.009), and found it to provide more active control (mean difference to modest 0.764; P=.009).

Figure 7. Interaction effects between information architecture and patient profile.



Discussion

Principal Findings

The aim of this study is to investigate how the organization, display, and structural design of a website, that is, IA, influences patients' experience with web-based patient education and the satisfaction and knowledge derived from the educational content. We wanted to understand whether user perceptions of engagement, control, personal relevance, trust, and novelty could explain how IA affects satisfaction and knowledge. Furthermore, we examined whether a user's profile affected which IA design was most effective or enjoyable to explore the potential of tailored IA design. Research on IA in the context of web-based health education has been sparse [10], which has limited intervention designers' ability to make informed design choices that enhance patients' experiences with web-based education.

This study compared three IA designs: tunnel, hierarchical, and matrix design. We found that in comparison with hierarchical IAs, tunnel and matrix IAs slightly improve user satisfaction. This effect may be explained by increased user perceptions of personal relevance in the tunnel IA and increased perceptions of control in the matrix IA. Contrary to our hypotheses and earlier findings [11], no direct or indirect effects of IA on knowledge acquisition or website use were found. However, the findings did indicate that IA preferences differ between patients with different user profiles. Specifically, patients with a so-called *managing* profile, who prefer open communication and have high communicative capabilities, are more satisfied with health education that is presented in a tunnel IA.

Our finding that tunnel IA design specifically affects satisfaction with emotional support is consistent with research showing that tunneled education improved the emotional well-being of patients with type 2 diabetes and chronic low back pain [60]. However, we did not replicate previous findings indicating that tunneling increases the use of web-based health interventions [11,22]. We did perceive a trend in this direction: participants in the tunnel condition used the website longer on average. However, this difference was not statistically significant. IA design did not predict knowledge acquisition either, despite previous findings indicating that tunneling improves knowledge acquisition [11]. Instead, user engagement emerged as the only predictor of knowledge acquisition. Some research on patient education indicates that cognitive factors such as working memory and cognitive load may be better predictors of knowledge acquisition than the user experience variables included in this study [61]. As IA design may facilitate cognitive processes, for example, by presenting information in smaller chunks as done in the hierarchical and tunnel designs, exploring whether IA design influences cognitive factors may be a worthwhile avenue for future studies that could help explain a larger portion of the variance in knowledge acquisition. In general, knowledge acquisition scores were low (47%-52%), which is in contrast with earlier findings that show that web-based patient education is effective for orthopedic patients [62] even when websites are consulted just once [63]. However, we are unsure whether these findings are due to inadequate education offered or poor source material (which was not changed when converted from paper to website) or because we did not test knowledge before the experiment. Regarding the latter, if participants had very little knowledge of TJR to begin with, it may be that although attained knowledge levels were low, they still represented decent knowledge acquisition. The participants' low self-reported knowledge of hip replacement supports this assumption: 81% of participants said that they had no or very limited prior knowledge. However, to fully answer this question, future research on IA design including pre-post measurements of health knowledge is needed.

The results of IA design on user engagement were mixed; the matrix IA achieved the highest subjective (ie, self-reported) engagement, but the tunnel IA was used the longest (albeit, not significantly longer). This reflects the dichotomous nature of engagement raised in the introduction, where engagement is thought to include both a subjective component of immersion and a behavioral component of use [15,64,65]. The findings indicate that IA design affects both but that matrix IA designs may be specifically suited for creating subjective experiences of engagement in patients. Furthermore, as only subjective self-engagement (and not duration of use) predicted actual knowledge scores, a very tentative conclusion may be that it might be more important to design engaging experiences rather than to design patient education materials that are used the longest. As most studies currently employ a *the more use, the better* perspective regarding engagement, use, and adherence to health interventions [66], this may require a different focus of researchers and designers alike.

Finally, this study focused on three simple IA designs for experimental clarity. Hybrid IAs that combine design elements from different IAs could mitigate the disadvantages associated with nonhybrid IAs. As users were most satisfied with matrix and tunnel IAs, hybrid matrix-tunnel designs should be explored further specifically. This study also identified that a large proportion of older adults with self-reported joint complaints use mobile phones (82%) and tablet devices (41%). As web-based IA designs cannot be ported to smartphones [13], IA designs suitable for health interventions distributed through mobile devices should be explored further. Finally, the field of IA has been affected considerably by the rise of recommender systems (RSs). These machine-based learning and information retrieval systems can predict and present relevant content, easing requirements for an adequate IA to help users locate content themselves. As this may diminish information overload [67], the potential benefits of combining RS techniques and IA in web-based health interventions warrant further research.

A secondary objective of this study is to explore the potential of tailored IAs. We found that participants with the highest information needs (so-called *managers*) preferred tunnel IAs. This finding supports the idea that patients' web-based learning experiences may be improved when IA is tailored to relevant user characteristics. However, we did not envision beforehand that the tunnel IA would actually match the *managing* profile. Rather, we assumed that participants in this group would prefer a matrix IA, as their skills, high self-efficacy, and preferences for openness and participation are in line with the theoretical *ideal* user of matrix IA websites [12]. According to the

qualitative feedback, one reason why they may have preferred the tunnel IA design instead is because it functioned as a checklist. Completeness or comprehensiveness is 1 of the 5 quality criteria for health information [68], and reassurance that all content had been covered may be particularly important for patients with high information needs. A tendency of patients with high information needs to actively seek out and ensure they have all available information (ie, a monitoring style of coping with threats) has been documented before in research with older patients with cancer [69]. Perceived comprehensiveness was not one of the mediators included in this study, but the question of whether some patients value it more than others, and which design elements may elicit comprehensiveness specifically, may be worthwhile avenues for future research. Finally, the patient profiles included in this study provided insight into orthopedic patients' skills and preferences for general communication, not digital communication, specifically. Effective use of eHealth requires composite skills beyond basic literacy, such as being able to operate search functions and knowing what information is available on the web [70]. Therefore, it may be more accurate to tailor IA design to eHealth literacy levels instead of a general profile.

In any case, the incongruence between anticipated and actual match of patient profile and IA design indicates that translating stated preferences to a tailored design is complex. Although the knowledge base on what works for whom is growing slowly, it may be more beneficial in the meantime to offer users a choice of IAs rather than dictating one design. Studies that explored the benefit of tailoring the mode of health information (eg, text, illustrations, audio-visual material) have successfully used *user-initiated tailoring* when working with multiple interfaces [71,72]. User-initiated tailoring requests users to customize a website's content and graphical user interface directly. Such customizations improve users' satisfaction, users' attention, and users' ability to recall knowledge [71,72]. Possibly, user-initiated tailoring may also be applicable to tailored IA design if users are offered a choice of IA designs when they first visit the website. A second consideration is to design IAs that support many different styles of health information processing. The work by Pang et al [73] on a website that was purposely designed to support 4 (rather than 1) distinct health information-seeking behaviors showed that users were more engaged with these dynamic interfaces. The communality between these studies is that users were not restricted or coerced to use the website in a particular way but instead were able to customize the experience to their self-determined preferences and needs at the time of visiting. Although this design approach may improve the fit between user and design, it may also introduce new issues (such as motivating people to adjust interfaces) that warrant further research. Yet, as more intricate eHealth interventions are developed and examined, it should be taken into consideration that these findings show that none of the examined IA designs had serious negative effects on satisfaction and knowledge acquisition and that although advantages in terms of improved user experience were present, they were small. The added value of highly customizable interventions should, therefore, be examined in tandem with the additional costs associated with developing multiple interfaces.

Strengths and Limitations

This study was conducted among adults who had self-reported joint complaints and may have viewed web-based education differently than patients scheduled for TJR surgery. However, previous studies have successfully tested health education websites in similar populations [11,71,72], and the high self-reported pain and lower health scores indicate that the study sample had considerable health concerns. As such, the sample can actually be considered a study strength as these individuals were likely motivated to learn about orthopedic health. At the same time, as the sample consisted solely of people with orthopedic health concerns, we know little of the generalizability of the findings of this study to other populations. Preferences for IA design may differ when using health education for purposes other than to prepare for TJR surgery (eg, to decide between alternative treatment options or to obtain support in managing a chronic illness), and additional research is needed to explore this.

Another limitation was self-selection, as participants were able to determine whether they wanted to join or leave the study. Between invitation for participation and inclusion in the study, 37% of participants were lost to follow-up. Of particular concern is that 6% of the sample dropped out after viewing the allocated website, as they might have done so based on their (negative) response to the website. This could make the study susceptible to type I errors [74,75]. This problem could not be remedied by intention-to-treat analysis due to the design of the experiment in which the participants that had dropped out generated no outcome data [75]. Therefore, we checked whether dropout was associated with allocation to a specific website, which was not the case. This made it unlikely that participants stopped because they were discontent with the allocated website. Another issue with self-selection was that participants could have been exceptionally interested in and already knowledgeable about TJR surgery. This would explain why we did not find any effects on knowledge. However, both self-reported prior knowledge of hip replacement and knowledge acquisition were generally low. A final limitation is that we determined satisfaction and knowledge gained from visiting the website once. As such, we cannot draw conclusions about experience with the website over time or knowledge retention after longer periods.

The strengths of this study include the experimental design. Although randomized experiments of website features known as *A/B tests* or web-based field experiments [76] are common in industry, the method is not often used in academic research on web-based health interventions. Various scholars have advocated moving beyond the *black box approach* which assesses only intervention efficacy. Testing specific features can help understand the mechanisms by which web-based interventions (do not) improve health outcomes [10,17,22]. By experimentally manipulating one feature and assessing both outcomes as well as mediating variables, this study takes a step in that direction. Second, the study took a human-centered and interdisciplinary approach to patient education design. The team included interaction designers, clinicians, and psychologists

and followed an iterative design process that involved patients early via pilot studies to ensure the usability of all three variants of the website. We believe that this commitment to developing three distinct but comparable, usable, and enjoyable web-based experiences has made it more likely that the effects on satisfaction can be attributed to differences in IA alone.

Conclusions and Recommendations for Intervention Design

Overall, our findings indicate that IA has small but notable effects on users' experiences with web-based health education interventions, at least in the context of orthopedic patient education. Tunnel IA design, in which users are guided through sequentially ordered content, improves perceptions of personal relevance and, in turn, user satisfaction. This design may be specifically appropriate for patients with high information needs. In contrast, providing users with more control over the way they progress through a web-based health intervention via a matrix IA design has positive effects on user perceptions of active control, which also contributes to higher satisfaction. Although additional research on IA design in different target groups and interventions is needed, hierarchical IA designs are not recommended at the moment, as hierarchical content is perceived as less supportive, engaging, and relevant, which may diminish the use and, in turn, the effect of the educational intervention.

Acknowledgments

This work is part of the research program Tailored health care through customer profiling (Project 314-99-118), which is financed by the Netherlands Organisation for Scientific Research (NWO) and Zimmer Biomet Inc. The authors thank the participants who took part in the pilot study and the consortium partners and reviewers for their constructive feedback on this work.

Conflicts of Interest

Part of the funding of this project is provided by Zimmer Biomet Inc (refer to the Acknowledgments section). This sponsor had no role in the study design of this protocol, data collection, analysis and interpretation, or writing of the report. In the case that this partner wants to apply for a patent based on research findings, publication can be postponed for a maximum of 3 months. No party has the right to prohibit the publication of these findings. The authors have full access to the study data.

Multimedia Appendix 1
Dutch translation of the User Engagement Scale-Short Form: validity, questionnaire items, and instructions for scoring.
[DOCX File , 33 KB - jmir_v23i3e15846_app1.docx]

Multimedia Appendix 2
Perceived advantages and disadvantages of tunnel, hierarchical, and matrix information architecture designs (translated from Dutch).
[DOCX File , 22 KB - jmir_v23i3e15846_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2224 KB - jmir_v23i3e15846_app3.pdf]

References

1. Langdon IJ, Hardin R, Learmonth ID. Informed consent for total hip arthroplasty: does a written information sheet improve recall by patients? Ann R Coll Surg Engl 2002 Nov;84(6):404-408 [FREE Full text] [doi: 10.1308/003588402760978201] [Medline: 12484580]
2. Turner P, Williams C. Informed consent: patients listen and read, but what information do they retain? N Z Med J 2002 Oct 25;115(1164):218. [Medline: 12552294]
3. Fagerlin A, Sepucha KR, Couper MP, Levin CA, Singer E, Zikmund-Fisher BJ. Patients' knowledge about 9 common health conditions: the DECISIONS survey. Med Decis Making 2010;30(5 Suppl):35-52. [doi: 10.1177/0272989X10378700] [Medline: 20881153]
4. Krupic F, Määttä S, Garellick G, Lyckhage ED, Kärrholm J. Preoperative information provided to Swedish and immigrant patients before total hip replacement. Med Arch 2012;66(6):399-404. [doi: 10.5455/medarh.2012.66.399-404] [Medline: 23409521]
5. Kinnersley P, Edwards A, Hood K, Cadbury N, Ryan R, Prout H, et al. Interventions before consultations for helping patients address their information needs. Cochrane Database Syst Rev 2007 Jul 18(3):CD004565. [doi: 10.1002/14651858.CD004565.pub2] [Medline: 17636767]
6. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]

7. Kebede MM, Liedtke TP, Möllers T, Pischke CR. Characterizing active ingredients of eHealth interventions targeting persons with poorly controlled type 2 Diabetes Mellitus using the behavior change techniques taxonomy: scoping review. J Med Internet Res 2017 Oct 12;19(10):348 [FREE Full text] [doi: 10.2196/jmir.7135] [Medline: 29025693]
8. Arden-Close EJ, Smith E, Bradbury K, Morrison L, Dennison L, Michaelides D, et al. A visualization tool to analyse usage of web-based interventions: the example of Positive Online Weight Reduction (POWeR). JMIR Hum Factors 2015 May 19;2(1):8 [FREE Full text] [doi: 10.2196/humanfactors.4310] [Medline: 27026372]
9. Morville P, Rosenfeld L. Information architecture for the World Wide Web. 2nd edition. In: O'Reilly Media. Sebastopol: O'Reilly Media; 2002:1600330150.
10. Pugatch J, Grenen E, Surla S, Schwarz M, Cole-Lewis H. Information architecture of web-based interventions to improve health outcomes: systematic review. J Med Internet Res 2018 Mar 21;20(3):97 [FREE Full text] [doi: 10.2196/jmir.7867] [Medline: 29563076]
11. Crutzen R, Cyr D, de Vries NK. The role of user control in adherence to and knowledge gained from a website: randomized comparison between a tunneled version and a freedom-of-choice version. J Med Internet Res 2012 Mar 09;14(2):45 [FREE Full text] [doi: 10.2196/jmir.1922] [Medline: 22532074]
12. Danaher BG, McKay HG, Seeley JR. The information architecture of behavior change websites. J Med Internet Res 2005 May 18;7(2):12 [FREE Full text] [doi: 10.2196/jmir.7.2.e12] [Medline: 15914459]
13. Danaher BG, Brendryen H, Seeley JR, Tyler MS, Woolley T. From black box to toolbox: outlining device functionality, engagement activities, and the pervasive information architecture of mHealth interventions. Internet Interv 2015 Mar 01;2(1):91-101 [FREE Full text] [doi: 10.1016/j.invent.2015.01.002] [Medline: 25750862]
14. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
15. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
16. Lynch P, Horton S. Web style guide: foundations of user experience design. 4th edition. New Haven, Ct: Yale University Press; 2016:0300211651.
17. Whitton AE, Proudfoot J, Clarke J, Birch M, Parker G, Manicavasagar V, et al. Breaking open the black box: isolating the most potent features of a web and mobile phone-based intervention for depression, anxiety, and stress. JMIR Ment Health 2015;2(1):3 [FREE Full text] [doi: 10.2196/mental.3573] [Medline: 26543909]
18. O'Brien H. Theoretical perspectives on user engagement. In: Why engagement matters. Switzerland: Springer International Publishing; 2016:1-26.
19. O'Brien HL, Cairns P, Hall M. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. Int Journal Human-Computer Studies 2018 Apr;112:28-39. [doi: 10.1016/j.ijhcs.2018.01.004]
20. Kelders SM, van Zyl LE, Ludden GDS. The concept and components of engagement in different domains applied to eHealth: a systematic scoping review. Front Psychol 2020 May 27;11:926 [FREE Full text] [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]
21. Ludden GD, van Rompay TJL, Kelders SM, van Gemert-Pijnen JEWC. How to increase reach and adherence of web-based interventions: a design research viewpoint. J Med Internet Res 2015 Jul 10;17(7):172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]
22. McClure JB, Shortreed SM, Bogart A, Derry H, Riggs K, St John J, et al. The effect of program design on engagement with an internet-based smoking intervention: randomized factorial trial. J Med Internet Res 2013 Mar 25;15(3):69 [FREE Full text] [doi: 10.2196/jmir.2508] [Medline: 23529377]
23. Liu Y. Developing a scale to measure the interactivity of websites. JAR 2003 Jun 01;43(2):207-216. [doi: 10.2501/JAR-43-2-207-216]
24. Liu Y, Shrum LJ. What is interactivity and is it always such a good thing? Implications of definition, person, and situation for the influence of interactivity on advertising effectiveness. J of Advert 2002 Dec;31(4):53-64. [doi: 10.1080/00913367.2002.10673685]
25. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
26. Song JH, Zinkhan GM. Determinants of perceived web site interactivity. J of Market 2008 Mar;72(2):99-113. [doi: 10.1509/jmkg.72.2.99]
27. Voorveld HAM, Neijens PC, Smit EG. The relation between actual and perceived interactivity. J of Advert 2011 Jul;40(2):77-92. [doi: 10.2753/joa0091-3367400206]
28. Lustria MLA, Cortese J, Gerend MA, Schmitt K, Kung YM, McLaughlin C. A model of tailoring effects: a randomized controlled trial examining the mechanisms of tailoring in a web-based STD screening intervention. Health Psychol 2016 Nov;35(11):1214-1224. [doi: 10.1037/hea0000399] [Medline: 27441869]

29. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008 Nov 04;10(5):36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
30. Kreuter MW, Strecher VJ, Glassman B. One size does not fit all: the case for tailoring print materials. Ann Behav Med 1999;21(4):276-283. [doi: 10.1007/BF02895958] [Medline: 10721433]
31. Dijkstra A. The psychology of tailoring-ingredients in computer-tailored persuasion. Social Pers Psych Compass 2008 Mar;2(2):765-784. [doi: 10.1111/j.1751-9004.2008.00081.x]
32. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
33. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
34. Yi MY, Yoon JJ, Davis JM, Lee T. Untangling the antecedents of initial trust in web-based health information: the roles of argument quality, source expertise, and user perceptions of information quality and risk. Decision Support Systems 2013 Apr;55(1):284-295. [doi: 10.1016/j.dss.2013.01.029]
35. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
36. Wathen CN, Burkell J. Believe it or not: factors influencing credibility on the web. J Am Soc Inf Sci 2002;53(2):134-144. [doi: 10.1002/asi.10016]
37. Rains SA, Karmikel CD. Health information-seeking and perceptions of website credibility: examining Web-use orientation, message characteristics, and structural features of websites. Computers in Human Behavior 2009 Mar;25(2):544-553. [doi: 10.1016/j.chb.2008.11.005]
38. Sillence E, Briggs P, Harris P, Fishwick L. A framework for understanding trust factors in web-based health advice. Int J Human-Computer Studies 2006 Aug;64(8):697-713 [FREE Full text] [doi: 10.1016/j.ijhcs.2006.02.007]
39. Huang M. Designing website attributes to induce experiential encounters. Computers in Human Behavior 2003 Jul;19(4):425-442. [doi: 10.1016/S0747-5632(02)00080-8]
40. Clark RE. Reconsidering research on learning from media. Review of Educational Research 2016 Jun 30;53(4):445-459. [doi: 10.3102/00346543053004445]
41. Groeneveld B, Melles M, Vehmeijer S, Mathijssen N, Dekkers T, Goossens R. Developing digital applications for tailored communication in orthopaedics using a research through design approach. Digit Health 2019;5:2055207618824919 [FREE Full text] [doi: 10.1177/2055207618824919] [Medline: 30719322]
42. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
43. Selya AS, Rose JS, Dierker LC, Hedeker D, Mermelstein RJ. A practical guide to calculating cohen's f(2), a measure of local effect size, from PROC MIXED. Front Psychol 2012;3:111 [FREE Full text] [doi: 10.3389/fpsyg.2012.00111] [Medline: 22529829]
44. Wiklund M, Kendler J, Strochlic A. Usability testing of medical devices. 2nd edition. In: Usability testing of medical devices. 2nd edition. Boca Raton, Florida: CRC Press; 2016:1-477.
45. Bol N, van Weert JCM, de Haes HCJM, Loos EF, de Heer S, Sikkel D, et al. Using cognitive and affective illustrations to enhance older adults' website satisfaction and recall of online cancer-related information. Health Commun 2014;29(7):678-688. [doi: 10.1080/10410236.2013.771560] [Medline: 24160560]
46. Bol N, Smets EMA, Rutgers MM, Burgers JA, de Haes HCJM, Loos EF, et al. Do videos improve website satisfaction and recall of online cancer-related information in older lung cancer patients? Patient Educ Couns 2013 Sep;92(3):404-412. [doi: 10.1016/j.pec.2013.06.004] [Medline: 23820196]
47. Jansen J, van Weert J, van der Meulen N, van Dulmen S, Heeren T, Bensing J. Recall in older cancer patients: measuring memory for medical information. Gerontologist 2008 Apr;48(2):149-157. [doi: 10.1093/geront/48.2.149] [Medline: 18483427]
48. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000 Dec 15;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014] [Medline: 11124735]
49. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract 2011 Apr;17(2):268-274. [doi: 10.1111/j.1365-2753.2010.01434.x] [Medline: 20874835]
50. O'Brien HL, Toms EG. The development and evaluation of a survey to measure user engagement. J Am Soc Inf Sci 2009 Oct 19;61(1):50-69. [doi: 10.1002/asi.21229]
51. Jensen JD, King AJ, Carcioppolo N, Davis L. Why are tailored messages more effective? A multiple mediation analysis of a breast cancer screening intervention. J Commun 2012 Oct;62(5):851-868 [FREE Full text] [doi: 10.1111/j.1460-2466.2012.01668.x] [Medline: 26405350]

52. Iacobucci D. Structural equations modeling: Fit Indices, sample size, and advanced topics. J Consumer Psychol 2010 Jan;20(1):90-98. [doi: 10.1016/j.jcps.2009.09.003]
53. Hooper D, Coughlan J, Mullen M. Structural equation modelling: guidelines for determining model fit. Electron J Bus Res Methods 2008;6(1):60. [doi: 10.21427/D7CF7R]
54. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.
55. Szende A, Williams A, editors. Measuring self-reported population health: an international perspective based on EQ-5D. Dordrecht Heidelberg New York London: Springer; 2004:1-115.
56. Essink-Bot ML, Stouthard ME, Bonsel GJ. Generalizability of valuations on health states collected with the EuroQolc-questionnaire. Health Econ 1993 Oct;2(3):237-246. [doi: 10.1002/hec.4730020307] [Medline: 8275169]
57. Dijkstra T. On statistical inference with parameter estimates on the boundary of the parameter space. Br J Math Stat Psychol 1992;45(2):309. [doi: 10.1111/j.2044-8317.1992.tb00994.x]
58. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986 Dec;51(6):1173-1182. [doi: 10.1037//0022-3514.51.6.1173] [Medline: 3806354]
59. Raykov T, Marcoulides GA. On desirability of parsimony in structural equation model selection. Structural Equation Modeling: A Multidisciplinary Journal 1999 Jan;6(3):292-300. [doi: 10.1080/10705519909540135]
60. Weymann N, Dirmaier J, von Wolff A, Kriston L, Härter M. Effectiveness of a web-based tailored interactive health communication application for patients with type 2 diabetes or chronic low back pain: randomized controlled trial. J Med Internet Res 2015 Mar 03;17(3):53 [FREE Full text] [doi: 10.2196/jmir.3904] [Medline: 25736340]
61. Wilson EAH, Wolf MS. Working memory and the design of health materials: a cognitive factors perspective. Patient Educ Couns 2009 Mar;74(3):318-322. [doi: 10.1016/j.pec.2008.11.005] [Medline: 19121915]
62. Dekkers T, Melles M, Groeneveld BS, de Ridder H. Web-based patient education in orthopedics: systematic review. J Med Internet Res 2018 Apr 23;20(4):143 [FREE Full text] [doi: 10.2196/jmir.9013] [Medline: 29685869]
63. Fraval A, Chandrananth J, Chong YM, Coventry LS, Tran P. Internet based patient education improves informed consent for elective orthopaedic surgery: a randomized controlled trial. BMC Musculoskelet Disord 2015 Feb 07;16:14 [FREE Full text] [doi: 10.1186/s12891-015-0466-9] [Medline: 25885962]
64. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
65. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
66. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
67. Cheung KL, Durusu D, Sui X, de Vries H. How recommender systems could support and enhance computer-tailored digital health programs: a scoping review. Digit Health 2019;5:2055207618824727 [FREE Full text] [doi: 10.1177/2055207618824727] [Medline: 30800414]
68. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. J Am Med Assoc 2002 May 22;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
69. Bol N, Linn AJ, Smets EM, Verdam MG, van Weert JC. Tailored communication for older patients with cancer: using cluster analysis to identify patient profiles based on information needs. J Geriatr Oncol 2020 Jul;11(6):944-950 [FREE Full text] [doi: 10.1016/j.jgo.2020.01.004] [Medline: 31974067]
70. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
71. Nguyen MH, Smets EMA, Bol N, Loos EF, Van Weert JCM. How tailoring the mode of information presentation influences younger and older adults' satisfaction with health websites. J Health Commun 2018;23(2):170-180. [doi: 10.1080/10810730.2017.1421729] [Medline: 29345531]
72. Nguyen MH, van Weert JCM, Bol N, Loos EF, Tytgat KMAJ, van de Ven AWH, et al. Tailoring the mode of information presentation: effects on younger and older adults' attention and recall of online information. Hum Commun Res 2016 Oct 21;43(1):102-126. [doi: 10.1111/hcre.12097]
73. Pang PC, Chang S, Verspoor K, Pearce J. Designing health websites based on users' web-based information-seeking behaviors: a mixed-method observational study. J Med Internet Res 2016 Jun 06;18(6):145 [FREE Full text] [doi: 10.2196/jmir.5661] [Medline: 27267955]
74. Wertz RT. Intention to treat: Once randomized, always analyzed. Clinical Aphasiology 1995;23:57-64 [FREE Full text]
75. Gupta SK. Intention-to-treat concept: a review. Perspect Clin Res 2011 Jul;2(3):109-112 [FREE Full text] [doi: 10.4103/2229-3485.83221] [Medline: 21897887]

76. Kohavi R, Longbotham R. Online controlled experiments and A/B testing. In: Sammut C, Webb GI, editors. Encyclopedia of Machine Learning and Data Mining. Boston, MA: Springer; 2017.

Abbreviations

ANOVA: analyses of variance
CFI: comparative fit index
IA: information architecture
MC: multiple choice
RMSEA: root mean square error of approximation
RS: recommender system
SRMR: standardized root mean square residual
THP: total hip prosthesis
TJR: total joint replacement
UES-SF: User Engagement Scale-Short Form

Edited by G Eysenbach, L Buis; submitted 12.08.19; peer-reviewed by G Ludden, N Vinson, MH Nguyen, C Chen; comments to author 15.05.20; revised version received 28.08.20; accepted 18.11.20; published 03.03.21.

Please cite as:
Dekkers T, Melles M, Vehmeijer SBW, de Ridder H
Effects of Information Architecture on the Effectiveness and User Experience of Web-Based Patient Education in Middle-Aged and Older Adults: Online Randomized Experiment
J Med Internet Res 2021;23(3):e15846
URL: https://www.jmir.org/2021/3/e15846
doi:10.2196/15846
PMID:33656446

Original Paper

User-Centered Development of a Web Platform Supporting Community-Based Health Care Organizations for Older Persons in Need of Support: Qualitative Focus Group Study

Verena Biehl[1,2], MA; Heidrun Becker[2], Prof Dr; Alenka Ogrin[3], MA; Alenka Reissner[3], PhD; Johannes Burger[4], MA; Andrea Glaessel[1,2,5], Prof Dr, MPH, MSc

[1]Institute of Health Sciences, Zurich University of Applied Sciences, Winterthur, Switzerland

[2]Research and Development in Occupational Therapy, Zurich University of Applied Sciences, Winterthur, Switzerland

[3]Zveza društev upokojencev Slovenije (ZDUS), Ljubljana, Slovenia

[4]SYNYO GmbH, Vienna, Austria

[5]Institute of Biomedical Ethics and History of Medicine, University of Zurich, Zurich, Switzerland

Corresponding Author:
Andrea Glaessel, Prof Dr, MPH, MSc
Institute of Health Sciences
Zurich University of Applied Sciences
Katharina-Sulzer-Platz 9
Winterthur, CH-8401
Switzerland
Phone: 41 58 934 4397
Email: andrea.glaessel@zhaw.ch

Abstract

Background: The ongoing changes in population demographics increase the relevance of dignified aging across Europe. Community-based health care (CBHC) organizations are necessary to provide sustainable strategies for organizing care for older persons in need of support. To support the digitalization of these organizations, new business models and suitable web platforms are necessary.

Objective: This study, which is part of the European Active and Assisted Living (AAL) project called “ICareCoops”, aimed to explore concepts, approaches, and workflows of CBHC organizations to achieve a comprehensive understanding of extant services offered and relevant requirements to support these services with information and computer technology (ICT) solutions.

Methods: A qualitative study with six focus groups (FGs) with 40 participants was conducted in Switzerland and Slovenia to identify potential stakeholders’ needs and requirements for the user-centered development of a web platform. Data were collected from three different stakeholder groups: (1) older persons in need of support as care receivers, (2) significant others of older persons in need of support, and (3) managers or care providers of CBHC organizations. A semistructured interview guide with open questions was used for data collection. FG sessions were audio-recorded and transcribed verbatim. Thematic content analysis was used to analyze the content of the FG sessions. To assist with further web platform development, the responses of the FG participants were translated into user stories to describe technical requirements.

Results: By analyzing the transcripts, five main categories were identified: (1) ICT usage behavior of users, (2) challenges of web platform usage, (3) content and technical requirements for the web platform, (4) form and services of CBHC organizations, and (5) rationales of CBHC organizations. The main issues identified were the need for seniors to have individual contact with the CBHC organization and the possibility to coordinate routine services via the web platform, such as ordering meals-on-wheels or booking a caregiver to accompany an older person to the doctor.

Conclusions: The majority of participants showed a lack of familiarity with the usage of ICT. Nevertheless, they were open-minded regarding web platform usage to facilitate workflows and to benefit CBHC organizations. Cooperatives as an organizational model demonstrate a high potential to address users’ needs. Therefore, the web platform offers an essential tool for innovative health care models in the future. Searching for care services, contacting care providers, and communicating with care providers was preferred via personal contact and seemed to be the key element for user acceptance and for the successful implementation of a web platform like “ICareCoops” to support CBHC organizations.

(J Med Internet Res 2021;23(3):e24006) doi:10.2196/24006

KEYWORDS

community-based health care services; older persons in need of support; user-centered design; focus groups; qualitative research; web platform

Introduction

Nowadays, older persons aged 65 years and older are still less familiar than younger generations with information and computer technology (ICT) usage. In Western societies, there is still a remarkable generation gap. In addition to age, other social and environmental determinants, such as reliable internet access, play a role in the use of ICT. Nevertheless, the percentage of ICT users among persons over 65 years has been rising in recent years [1-3]. Older persons are more likely to need support in daily life to stay at home independently because of morbidity, immobility, or general frailty. In the coming years, the proportion of the population aged 60 years and older will increase worldwide, specifically one in five people will be aged 60 years or older in 2050 [4]. Because of this demographic development, dignified and healthy aging and caregiving is becoming a more relevant topic in the long-term perspective as European health care systems are facing challenges of sustainability and an increasing need of resources [4,5]. As a consequence of these demographic changes, not only will the demand for health care services for older persons in need of support increase but a coordinated cooperation of services will be necessary to maintain seniors' autonomous living at home [4,6]. In particular, ICT has a growing potential to provide essential assistance for older persons in need of support in finding and coordinating individual requirements for assistance with living and health care services. These ICT tools still have to be developed and implemented in a way that is consistently user-friendly yet adaptable for evolving health care models [6-8].

Besides public health care services, more and more nonprofit organizations are becoming major players in delivering health care services in communities [9-11]. Nonprofit organizations are based on five criteria, which are (1) organized, (2) private, (3) not profit-distributing, (4) self-governing, and (5) noncompulsory [10]. These mainly community-based organizations deliver a broad spectrum of social and health-related services, such as providing informal counselling and social support, building individual and community capacity, coordinating health services, and supporting self-management in older persons in need of support [12]. A well-known example of a nonprofit organizational form, which also comprises community-based health care (CBHC) services, are care cooperatives. The difference between cooperatives and other forms of enterprises is that profit-making and economic stability are balanced by the interests of the community [13]. Cooperatives are autonomous associations of people who voluntarily cooperate for their mutual social, economic, regional, and cultural benefit. Cooperatives balance multiple stakeholders' interests instead of shareholders' interests alone [9]. Cooperatives as a person-centered business model can become health care actors with the potential for large cost savings, service improvement, and more sustainable cooperation between those responsible for health care.

To date, care cooperatives in Europe have been established mainly in Italy, Turkey, France, and Spain [14]. In Switzerland, there are some care cooperatives (eg, Spitex, a Swiss nonprofit outpatient care cooperative [15]) and medical centers in rural areas. In the past, some of the cooperatives developed into good practice examples, whereas others failed because of poor communication, coordination, etc. Especially for these reasons, ICT tools could have a large impact on improving their services, addressing not only the concrete needs of older persons in need of support but also the daily operations of such organizations. Although the care cooperative model seems to be highly suitable for CBHC services, in this study we did not focus on one specific organizational form of CBHC service in the nonprofit sector. Instead, we focused on different CBHC organizations to identify the potential benefits of ICT support in their daily operations. Web platforms are a useful tool, but to the authors' knowledge they have not yet been broadly implemented in CBHC organizations.

This study is part of a larger project, the European Active and Assisted Living (AAL) project called "ICareCoops", that is targeted at CBHC organizations and aims to develop a web platform [16] for implementing and managing CBHC services for older persons in need of support. From the beginning, high user acceptance of the web platform has been set as the project's focus. Therefore, a user-centered approach was chosen, which is characterized by heavy involvement of future web platform users. This user-centered design optimizes future ICT user participation, and it is a common method to develop digital health-related tools [17,18]. This study is one part of the web platform development.

The objective of this study was to explore concepts, approaches, and workflow processes of CBHC organizations and their potential as care cooperatives. Specifically, we pursued a two-part goal: (1) to achieve a comprehensive understanding of extant services offered and relevant requirements including risks and challenges of using ICT, and (2) to support these services with ICT solutions, exemplified in two European countries (Switzerland and Slovenia).

Methods

Study Design

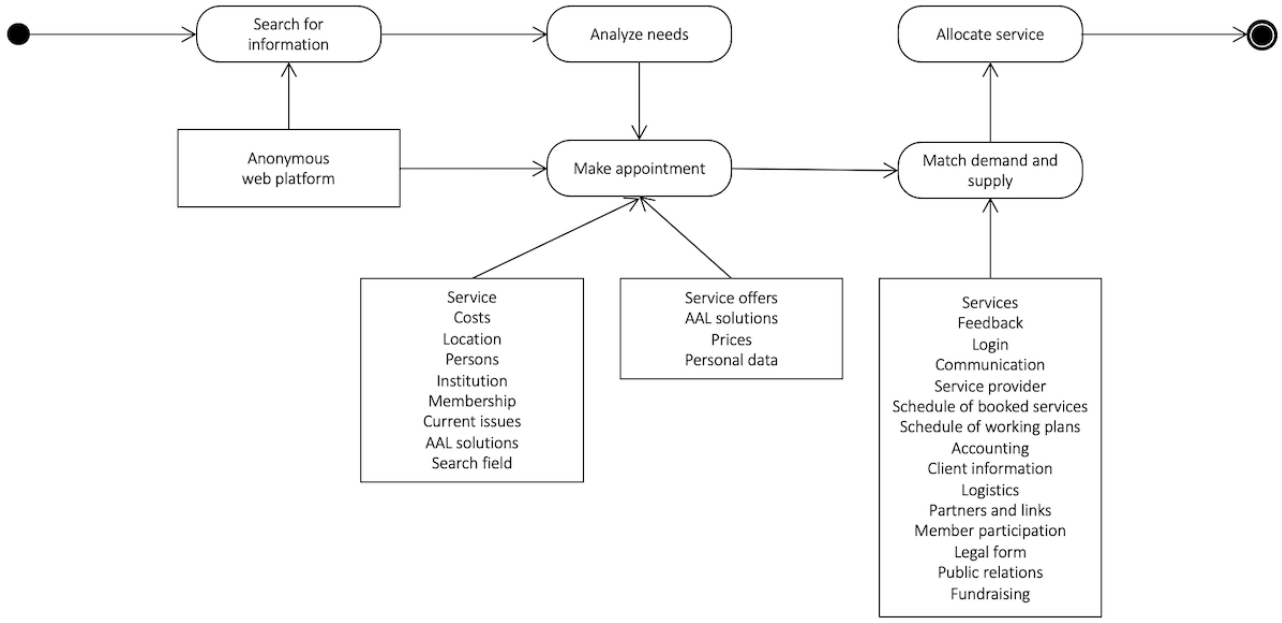
To address the study's goal, a user-centered design based on participatory research was chosen. Moreover, by using focus groups (FGs), a qualitative methodology was applied to identify the most relevant requirements from the users' perspective. FGs were useful to identify the needs of target groups and to gather a broad spectrum of information and opinions from older persons in need of support. In our case, stakeholders comprised the following three groups: (1) older persons in need of support as

care receivers of a CBHC service, (2) significant others of older persons in need of support, and (3) managers or care providers of a CBHC service in Switzerland or Slovenia. Discussing issues with multiple participants of the same stakeholder group reveals a broader variety of relevant aspects and useful ideas [19]. All participants were informed about the project and provided consent to participate in this study based on written and spoken information. FGs were treated with full confidentiality, and anonymity was maintained throughout the research process.

In qualitative studies, the role and reflexivity of the researchers is an influencing factor during the entire research process. Actions, perceptions, and especially context-relevant aspects are important and considered in the studied field—in this study, future users of web-based CBHC services for older persons in need of support. In qualitative research, the subjective perceptions of the researchers and users as a component of knowledge are applied to understand complex contexts [20].

To address all objectives within this qualitative study, we chose a mix of deductive and inductive approaches. The deductive component consisted of applying a workflow model (Figure 1) to demonstrate to the FG participants a potential template of a web platform and to discuss the different processes of a workflow within a CBHC organization in depth. Illustrating this simplified workflow model should enable tech-novice FG participants to discuss potential workflows that could be supported with ICT tools. To comply with the user-centered approach, the workflow model intentionally used everyday speech and its design was kept simple. In addition, the inductive component, with open-ended questions, facilitated gathering further information about FG participants. Detailed information on the data collection process is described further below under the “Data Collection” subsection.

Figure 1. Workflow model of a community-based health care organization, including potential tools for support from a web platform. AAL: Active and Assisted Living.



Sampling

For each FG, approximately 9 participants were invited by email and phone to join the FG to represent the needs and expectations of all three stakeholder groups. We considered differences in the requirements between the two countries, in the characteristics of the health care organizations (eg, year of foundation, size, ICT use), and between different stakeholders (older persons in need of support and their significant others and managers/care providers). To avoid the predominance of input from professionals, as well as mixing up knowledge and expectations of managers/care providers and older persons in need of support, the web platform requirements were discussed separately from three perspectives. Among the stakeholders, a heterogeneous group composition of participants was attained, allowing the collection of richer information [19].

Recruitment Strategy

Participants from different economic and cultural backgrounds were consecutively selected in Slovenia and Switzerland. Swiss participants were recruited from the German-speaking region of Switzerland. In Slovenia, organizations located around the city of Ljubljana were contacted. In both countries, potential participating CBHC organizations received information about the project and were requested to distribute the invitation to their members. Furthermore, we asked them to send managers from an institution. We also ensured to include a variety of technologically competent people. The intention was to gain a better insight into the variety of workflows and requirements needed for the envisaged web platform.

Ethical Approval

In Switzerland and Slovenia, ethical approval is required for clinical trials—for extraction of biological material or collection of health-related data [21,22]. This project is primarily concerned with the technical development of a software solution.

End users were invited to contribute to the process of requirements engineering of software solutions. Given the nonmedical context, ethical approval was not required for all involved countries. Collected data were used exclusively for generating software requirements and improving the prototype. No risks or damages were foreseen during the end user FG study. Personal data of the end users was anonymized, codified, and stored in a secure place, guaranteeing access only to authorized persons and safeguarding the right to privacy. The software prototype does not contain any personal information from end users. Informed consent highlighted the possibility of research results being published in scientific journals or being presented at conferences, always with the guarantee of anonymity. All participants had an exit right and could withdraw from the project at any time without giving a reason.

Data Collection

To gather sociodemographic data from our sample, participants were asked to complete a questionnaire. During the FG process, we followed a semistructured topic guide with four open-ended questions. All four questions were based on scientific literature searches, including best-practice models of workflows and web platforms of CBHC organizations.

For this study, we developed a simplified workflow model as a basis for discussion with the stakeholders (Figure 1). This model comprised the most relevant four steps within the workflow of a CBHC organization and expected stakeholder requirements. Subsequently, we tried to combine the consecutive workflows within the potential web platform of a CBHC organization:

1. Need for care support: from a user perspective, the workflow starts with an internet search.
2. Information: relevant information has to be presented on the web platform concerning the organization (eg, name, membership, region, services, etc).
3. Needs analysis: the workflow continues by contacting the CBHC organization and making an appointment to assess the demands of the user. In this step, we wanted to know which kinds of ICT users would use (eg, email, telephone, video call, apps, etc).
4. Allocation of services: after an assessment of needs, services had to be selected and offered.

The model considered the three different stakeholder groups. For all three target groups, needs for working and cooperating in a CBHC organization with a variety of requirements (eg, price checking, membership management, scheduling, communicating with others, etc) were discussed.

This workflow model was one component of the FG topic guide. Furthermore, the guide included an introductory section about the project, study-specific open-ended questions pertaining to users' needs and expectations, time for further issues, and an outlook of the following project steps.

Study-specific open-ended questions used to derive user-specific requirements, expectations, and challenges related to ICT-based support of a CBHC organization are illustrated in Table 1.

Table 1. Open-ended questions in the topic guide for focus groups from three perspectives.[a]

Topic	Questions
Identify workflows including ICT[b] solutions	• Which other services would be useful in your organization? • If you think long term, which services could be relevant in the future?
Identify desirable ICT solutions	• Which ICT supports do you think would be helpful to care organizations? • Which functions should be possible?
Identify barriers to ICT usage	• Which barriers/risks do you identify in the usage of ICT solutions? (eg, cultural, ethical, legal—personal and general)
Identify potential of a care cooperative	• Name the 2 to 3 most relevant advantages of founding a care cooperative from your point of view.

[a]Perspectives of older persons in need of support as care receiver, significant others of care receivers, and managers or care providers of community-based health care organizations.

[b]ICT: information and computer technology.

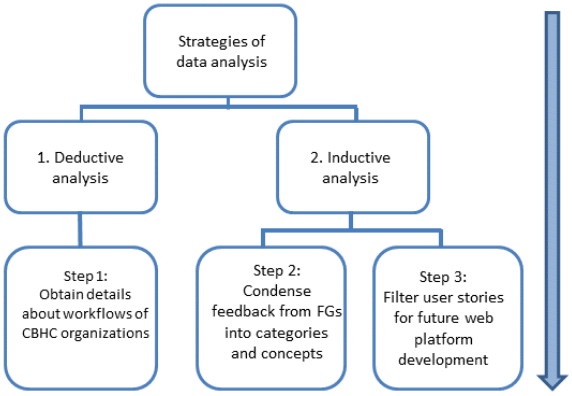
Four researchers conducted the FGs—two in Switzerland and two in Slovenia. One researcher acted as moderator and one assisted by taking field notes. To guide the discussions and focus on the topic, questions were presented as PowerPoint slides. A duration of 2 hours for one session was not exceeded. Each FG session was audio-recorded and transcribed verbatim after receiving written consent from the participants. In all transcripts, protocols, and case reports, names and personal data were anonymized.

Data Analysis

Data were extracted by thematic content analysis based on written transcripts, which was conducted in three steps, characterized by an inductive and deductive strategy to receive two different kinds of data (Figure 2). In step one, the deductive data analysis, we aimed to obtain details about the workflows of CBHC organizations. Statements and responses from FG participants were summarized according to the topics in the workflow model. In step two, the inductive data analysis, information, new themes, and issues mentioned by FG participants were first read in-depth, open codes were generated,

and the data were condensed into concepts and categories. Microsoft Excel 2010 was used for the structured data analysis. In step three, user platform requirements had to be filtered systematically for further web platform development. Based on the results of steps 1 and 2, technically relevant results were translated into user stories of the syntax "<as an> <I want to> <so that>." The user stories will be the basis for further technical web platform development in the future and were not part of this study, which focuses on an analysis of specific needs of potential web platform users. User stories are a method that was applied to gather user requirements for developing an agile web platform [23].

Figure 2. Strategies and steps of focus group data analysis. CBHC: community-based health care; FG: focus group.



Quality Assurance and Trustworthiness

Data security regulations regarding data storage were observed throughout the study. For data quality assurance, the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist was applied [24]. For peer review, a pretest of the FG questions in the topic guide was conducted. Field notes were collected by a second researcher. After each FG session, both researchers reflected on the process for potential improvements with a written debriefing. For peer review, a second researcher coded 25% of the data from the transcripts. The results were discussed in order to reach consensus on each concept. All participants of the study received feedback in the form of a plain-language summary of the study results via email.

Results

FG Sessions

In Switzerland and Slovenia, six FG sessions, each with a maximal duration of 2 hours, were conducted with three target groups: (1) older persons in need of support as care receivers of a CBHC service, (2) significant others of older persons in need of a CBHC service, and (3) managers or care providers of a CBHC service. Of the 10 organizations who were invited, 70% (n=7) participated in the FGs by sending different stakeholders: managers, care providers, older persons in need of support, and significant others of older persons in need of support.

Sociodemographic Data

In total, six FGs with 40 participants (32 women, 8 men) were conducted. Personal characteristics indicated a large variety of tech-savvy and non–tech-savvy people and formal and informal care providers and receivers and their significant others. Sociodemographic data of FG participants are presented in Table 2.

Table 2. Sociodemographic data from the three target groups in the focus groups.

Country of data collection and target group	Participants, n	Females, n (%)	Age (years), median (range)	Characteristics of sample
Switzerland				
Older persons in need of support	7	5 (71)	75 (67-83)	Tech-savvy and non–tech-savvy; formal and informal care receivers; potential care receivers; members of CBHC organizations and non-members
Significant others of older persons in need of support	7	4 (57)	64 (53-68)	Tech-savvy; informal care providers and potential informal care providers for family members
Manager/care provider	8	5 (63)	71 (65-80)	Tech-savvy and non–tech-savvy; informal care providers; management members of CBHC organizations
Total	22	14 (64)	70 (53-83)	
Slovenia				
Older persons in need of support	4	4 (100)	72 (69-83)	Non–tech-savvy; basic education; low-income; formal care receivers
Significant others of older persons in need of support	7	7 (100)	37 (28-58)	Tech-savvy; informal care providers and potential informal care providers for family members
Manager/care provider	7	7 (100)	65 (59-79)	Tech-savvy and non–tech-savvy; informal care providers; management members of CBHC organizations
Total	18	18 (100)	64 (28-83)	
Switzerland and Slovenia (total)	40	32 (80)	67 (28-83)	

Deductive Analysis of Organizations' Workflows

Participants from the manager/care provider group mainly agreed on the workflow model (Figure 1). Usually, older persons in need of support call the organization by phone and ask for services. In most cases, service providers initiate a personal appointment with the service receiver to identify needs to establish a personal, face-to-face relationship with the older person in need of support.

> *To our understanding the first contact has to be free of charge.* [Relative of care receiver 03_Slo]
>
> *Personal contact is central to all of our members. This will remain the most important rationale of CBHC organization beyond the next ten years.* [Care provider 04_Swiss]

After determining the patient's needs, the organization arranges tailored services for the older person in need of support. The FG participants recommended involving older persons in need of support and their significant others in the decision-making process for care services. A test period of services would be beneficial for older persons in need of support. After this period, they could decide whether to continue with the services or not. Participants recommended that a contract should be signed between older persons in need of support and CBHC organizations to clarify arrangements, frequency of services, and payment terms. In most organizations, a person in the role of a distributor works in the office. Service users wished to have a contact person in the organization who could coordinate the requests and service offers. In a CBHC organization, a distributor is required to monitor services, contact older persons in need of support, and coordinate formal and informal care providers. Moreover, the same person determines conditions and deadlines of payments according to user needs and facilitates communication among all stakeholders. A participant in the provider FG in Switzerland emphasized, "The organization would not work without a contact person who coordinates all the stakeholders' needs" [Care provider 02_Swiss].

Inductive Analysis

In the inductive analysis, five main categories were extracted from open coding of data from the three stakeholder groups: (1) ICT usage behavior of users, (2) challenges of web platform usage, (3) content and technical requirements for the web platform, (4) form and services of CBHC organizations, and (5) rationales of CBHC organizations. Each category is described in the following sections, including examples of codes. The results of categories 1, 4, and 5 are illustrated in Table 3. The results of categories 2 and 3 are presented separately in Table 4 and Multimedia Appendix 1, respectively.

Table 3. Subcategories and codes of three of the five main categories extracted from the study data.

Category and subcategories	Codes
ICT[a] usage behavior	
Current ICT usage among older persons	• ICT usage rises in the next 10-20 years • No idea about options of ICT • Significant others more familiar with ICT usage • Small proportion of the elderly is familiar with ICT • Open-minded attitude regarding ICT usage • No information accessible without the internet
Current ICT usage among CBHC[b] organizations	• Use of ICT via phones, Excel, and voicemails • Computer cannot substitute a person or individualize the services • Coordination of services is performed by people (ie, not possible without them) • ICT usage by managers of organizations, only by a few members
Usage of web platform	• Benefit greater for managers of CBHC organizations • Participants would be interested in using the web platform • Future generations to benefit more
Form and services of CBHC organizations	
Organizational form: association or cooperative	• Associations as business models are more represented than cooperatives • Liability problems are excluded in associations • Associations allow spontaneous freedom of action • No difference between associations and cooperatives
Current and desirable services of CBHC organizations	• Independent counselling without the intention of selling something • Carpools • General practitioners and health professionals in care cooperatives • Deposit of private data, such as patient's provision or testaments • Individual services adjusted to available budget • Driverless vehicles (AAL[c] project) • Rental robots (AAL project) • High-quality meals-on-wheels • Support in emergency situations (eg, correct phone numbers, weekend assistance) • Coordination of professional care • Expansion and optimization of palliative care
Possible concepts for the work of CBHC organizations	• Simple neighborhood aid • Mixed generations (mutual aid) • Living concepts for mixed generations • Older persons as informal caregivers (mutual aid) • Refugees as informal caregivers
Rationales of CBHC organizations	
Rationales together with older persons in need of support	• Inhibition in asking for support (neighbors, family) • Concerns regarding nursing homes • Physical and mental limitations normal • Demand required for internet usage • Nonacceptance of support • Societal changes, disruption of family structures • Personal contact very important (prevent social isolation) • Autonomy as long as possible
Rationales collaborating with care receivers/members of the organization	• Never say no to member requests • Social contact is necessary for needs assessment • Need to talk to someone, as social contacts decrease • Personal conversations essential • Care requires consideration of logistics (coordination is important) • Social contacts arise spontaneously

Category and subcategories	Codes
Rationales cooperating with volunteers	• No bureaucratic effort, less scheduling • Motivated volunteers are available • No exploitation of volunteers • Recompense for volunteers (eg, time credit, tax reduction)

[a]ICT: information and computer technology.

[b]CBHC: community-based health care.

[c]AAL: Active and Assisted Living.

Table 4. Subcategories and codes of main category 2: challenges of web platform usage.

Subcategories	Codes
Common issues	• Information overload on the internet (or on the web platform, respectively) • Transmission time of the internet can be slow • The knowledge gap in using the internet has to be filled in the target group • Technical products (AAL[a] project) and requirements change very quickly nowadays
Ethical issues	• Protection of sensitive data has to be respected • User awareness that the internet works in the same way as a monitoring system has to be raised • Concerns about moneymaking with the web platform • Health concerns have to be respected (people with dementia cannot use the platform and book services)
Financial issues	• Not everyone can afford internet access • Not everyone can afford the required hardware (smartphone, PC, tablet, AAL) • Retired people cannot afford any additional expenditures (services, hardware) • There has to be a wide range of payment options (credit card, PayPal, bank transfer) • Risk for seniors of buying/ordering too much without control
Cultural issues	• Not everyone has access to the internet (especially this target group) • Some people refuse to use the internet/TV, etc • Target group is afraid of the future regarding internet usage (and its possibilities) • Owning a smartphone is more common than owning a computer (ie, apps would be more helpful than PCs)

[a]AAL: Active and Assisted Living.

ICT Usage Behavior of Users

In this category, the target groups' current ICT usage behavior was compiled. All statements concerning the issue of ICT usage of seniors supplement this category.

Researchers recognized a great variance in tech-savviness among participants. The user (older persons in need of support as care receivers) FG in Switzerland was more tech-savvy than managers or care providers. In contrast, the user group in Slovenia did not have any experience in using ICT. Thus, all FGs were conducted on a very different level of understanding regarding the development of a web platform. Overall, participants assumed that modern ICT cannot be used by approximately 50% to 80% of older persons in need of support because of lacking skills or familiarity with ICT usage. At present, only a small sample of tech-savvy seniors would benefit from such a web platform, although future generations may benefit from the web platform. A period of about 10 to 20 years is predicted for older persons in need of support to become sufficiently tech-savvy to use a web platform.

Older persons in need of support would be interested in improving their skills and therefore suggested courses as a requirement for establishing this web platform. They also assumed that care providers would benefit more from such a web platform from a management and organizational point of view.

The most common reported uses of ICT were phone calls for making appointments, emails for confirmation of appointment dates, and Excel spreadsheets to collect information about care receivers/members/providers. CBHC organizations knew that current ICT usage is very poor, but they did not have resources to invest in this issue and were satisfied with their current state. Most of the FG participants had difficulties in anticipating future needs and providing innovative ICT solutions; however, they all demonstrated an openness to using such a web platform if it was user-friendly and easy to access and navigate.

Challenges of Web Platform Usage

Participants were asked to discuss challenges of using the web platform. The identified barriers could be divided into four subcategories: (1) common issues, (2) ethical issues, (3) financial issues, and (4) cultural issues (Table 4). Low ICT usage among the target group of seniors was seen as a main challenge for the web platform; they would assume there to be an information overload on the internet and would therefore only use an easy and clearly arranged web platform.

> *If I go online and search, for example, meals-on-wheels, I will find so much information and*

> *I can't handle this information. I need only some high-quality results in the region where I live. If the web platform could filter this information for my needs and was easy to use and clearly structured, I would appreciate this web platform.* [Care receiver 04_Swiss]

Furthermore, trust in data protection among the target group was critical. They reported being afraid of storing sensitive data on a web platform.

Content and Technical Requirements for the Web Platform

As a main result of FG discussions, specific requirements for the web platform were mentioned and could be retrieved. The requirements covered common, technical, and content aspects of the web platform in regard to hardware and mobile apps; these requirements are summarized in Multimedia Appendix 1.

> *The ICT system should be very well built, and someone should monitor it and organize the work. For the care-providing organization, it is important to have good visibility and access to data.* [Care provider 03_Slovenia]

All of the participants confirmed that there need to be people in the background to organize and manage all of the services and member requests; one participant in Switzerland called it an "intelligent system" [Relative of care receiver 01_Swiss]. All participants were quite open-minded regarding hardware and potential software solutions offered. They mentioned possibilities such as fingerprint identification, voice-guided tools, video calls, and video tutorials to explain the web platform. In total, participants named 60 requirements, which were then translated into user stories for further web platform development in a separate step of the project. The separate step is not part of this study. Examples of user stories include the following:

> *I as a care provider have the option to contact the care receiver via an app, so that it's possible to inform about changes in the schedule, etc.*
>
> *I as a care receiver have video tutorials which explain important information (service), so that I understand and get an idea of the services offered or other important information (how can I become a member, how can I offer help…).*

Overall, the idea of CBHC organizations working with such a web platform was very well received, with the prerequisite that it would be user-friendly and accessible without a high cost.

Form and Services of CBHC Organizations

This category includes information concerning the business model of a CBHC organization comparing associations and cooperatives. The main task of a CBHC organization as a service provider is to fulfil the target group's needs, which are also described by this category. For care providers, the organizational model of cooperatives is currently not widespread in Switzerland. A widely used business model for existing institutions is an association because it is easier to administer (eg, regarding liability of partners). Differences between acting as an association and acting as a cooperative in health care were not obvious to FG participants.

> *It is easier to work as an association in Switzerland than as a cooperative as legal form for various reasons.* [Care provider 01_Swiss]

Nevertheless, the motivation to work with cooperative principles is untouched by this solution. Slovenian participants emphasized that they did not have cooperatives in Slovenia, but they assumed it was an engaging idea. In general, the suggestion of care cooperatives was convincing to participants in both countries.

The variety of current and desirable service offers in CBHC organizations is very broad and ranges from collective buying (eg, of medications) to assistance in housekeeping. Organizations offer an endless number of different services, except professional care. They do not offer health or nursing care because this is provided by special health care organizations. Participants mentioned the desire to simplify health care by including professional care and practitioners to coordinate all care services offered by CBHC organizations. To give an impression of services offered, care organizations brought information materials to the group sessions; their services included mowing the lawn, accompanying someone to the doctor, transportation services, and special training for seniors, such as computer courses or universities for seniors. It was also identified that seniors can rarely afford additional expenditures, such as care services.

> *Therefore, the services have to be affordable even to people with low incomes; usually they are co-financed by municipalities.* [Care provider 01_Slovenia]

Moreover, dissatisfaction with professional care in health care organizations was revealed. Primarily for potential users of CBHC organizations, some services should be strengthened, such as high-quality meals-on-wheels, emergency aid at night and on weekends, and regional information for seniors convened at one place. In general, participants seemed to have a need for information about any issues related to seniors located in one place (eg, care, courses, communication, leisure, etc).

> *The CBHC organization could act as an information center.* [Relatives of care receiver 03_Swiss]

CBHC organizations should (1) inform users via different channels (internet, telephone, in person, printed documents) about any issues related to seniors; (2) provide customized recommendations for care needs; and (3) store data about users and transfer them to relevant stakeholders, such as care providers or practitioners. Furthermore, ideas about AAL solutions were suggested, such as robots to support housekeeping or fully automated, driverless vehicles.

Participants discussed possible concepts for CBHC organizations to involve retired people in the informal care of seniors as well as intergenerational projects. Younger generations could assist older persons in need of support, and older persons in need of support could assist families in other areas of need (eg, babysitting). Furthermore, living concepts, where seniors live with younger generations in the same housing facility, were mentioned as desirable.

Rationales of CBHC Organizations

This category summarizes statements of the target groups about fundamental principles to be considered in CBHC organizations, as they seem to be of particular importance for collaboration with older persons in need of support.

Participants mentioned rationales for the collaboration of CBHC organizations with older persons in need of support; participants reported that autonomy was desirable if possible. Therefore, a decrease of physical and mental mobility is seen as an obstacle, as it could potentially lead to a loss of autonomy. Nowadays, family structures are supposed to be less stable and close so that older persons in need of support do not require support from the family.

> *I don't want to disturb my children. They are very busy with a lot of other things.* [Care receiver 06_Swiss]

Beyond that, older persons in need of support worry about moving to nursing homes. Here one can see a discrepancy between the intention to live independently, the fear of requesting support or moving to a nursing home, and the need for support caused by a decrease in physical and mental mobility. Participants confirmed that they needed support when using the internet; they would probably need supervision when using ICT.

Nearly all FG participants ranked having a personal contact within a CBHC organization as the most important feature. Older persons in need of support need to have a contact person within the organization that they refer to. They assumed social isolation would be a risk factor as ICT use increases.

> *It is not an exception that service providers and mediators of the informal care organizations are the only contact persons that seniors still have.* [Care provider 01_Swiss]

Booking "social contact" as a service offering is perceived as taboo. For many, social contacts are to be established spontaneously and not by booking an appointment. Participants gave a clear definition of when personal contact to organizations is required: as soon as the care service becomes intimate (ie, someone coming into their house or taking over intimate care services like hygiene).

A further principle of all participating care organizations reported by participants is

> *… to provide as much support as possible to assist members. Usually we don't say NO to our members.* [Care provider 08_Swiss]

CBHC organizations tend to decline none of the requests of their members. In their daily business, they feel challenged and responsible for creating new solutions, also for nontypical and nonstandard needs.

Participants confirmed that there are volunteers motivated to act as informal caregivers, but efficient ways of collaboration need to be established. This could be achieved by paying overtime or monetary credits for volunteers. Collaboration should work with small bureaucratic effort for volunteers. Nonfeasible expenditure/investment would exclude this important group of informal care givers.

Discussion

Principal Findings

This qualitative FG study aimed to explore concepts, approaches, and workflows of CBHC organizations in order to achieve a comprehensive understanding of existing services offered and relevant requirements to support these services with ICT solutions exemplified in two European countries. The Swiss and Slovenian results are related to three stakeholder groups: (1) older persons in need of support as care receivers of a CBHC service, (2) significant others of older persons in need of support, and (3) managers or care providers of a CBHC service. Results from the deductive analysis of existing workflows of CBHC organizations and the potential need of ICT support in the different workflows will be discussed, as well as the results from the inductive analysis based on five main categories: (1) ICT usage behavior of users, (2) challenges of web platform usage, (3) content and technical requirements for the web platform, (4) form and services of CBHC organizations, and (5) rationales of CBHC organizations.

To conclude, the idea of CBHC organizations delivering formal and informal care is very popular and is recommended among stakeholders. Participants emphasized personal contact between members and CBHC organizations as a main rationale for CBHC organizations. Social contact is seen as highly important to older persons in need of support. The main challenge when using a web platform such as the one being proposed will be the knowledge gap in ICT usage among older persons in need of support. Therefore, an age-appropriate web platform design and seniors' mistrust in data protection need to be addressed. ICT usage will likely rise during the COVID-19 pandemic, but it can also lead to even larger "digital divides" among older persons in need of support.

Results of technical and content-based requirements for the web platform showed the openness of potential users regarding hardware and software solutions. FG participants emphasized the importance of an "intelligent system" behind the web platform in the form of a person or organization that coordinates and leads the web platform. The workflow of a CBHC organization proposed by the project group (Figure 1) was largely acknowledged by FG participants. Furthermore, it can be concluded that stakeholder results between different FGs were widely congruent. Certainly, there were differences in the aspects discussed, as there were three different stakeholder perspectives, but they shared the same ideas about principles of CBHC organizations, challenges of ICT usage, and the content and technical requirements for the web platform. Major differences in the issues mentioned by the different FGs were noted in the results section and will be discussed further below.

ICT Usage and its Challenges

A main issue in the discussions about the web platform was the lack of knowledge of ICTs and the habits and skills of older persons in need of support with ICT applications. Apparently, older persons in need of support as care receivers and managers

or care providers of CBHC organizations could not imagine the possibilities of ICT. One of the reasons why a third stakeholder perspective—of the significant others of care receivers—was needed in this study was because the younger generation is expected to be more tech-savvy and will profit more from such a web platform. In Switzerland, older persons in need of support as care receivers confirmed that many of them were quite familiar with using ICT daily for information seeking and communication with family and friends. It was easier for them to imagine a web platform that contains tools to support CBHC organizations. A reason for high tech-savviness of Swiss older persons in need of support could be that those interested in using ICT may follow an invitation to a discussion on this topic rather than care receivers who are unfamiliar with ICT. In Slovenia, in contrast, the older persons in need of support were not familiar with ICT usage. This helps to gain a broad understanding of the variability of ICT usage among this target group in Europe. Daily internet usage in Switzerland is slightly above the European Union average [1].

Managers and care providers insisted on their current concepts and workflows in their CBHC organizations and could not imagine a web platform supporting and coordinating members and care providers. Older persons in need of support were aware of their knowledge gap and lack of practice in using ICT. They were convinced that future generations would benefit from this web platform. Moreover, they were open-minded and interested in taking ICT courses to improve their skills. These results are confirmed by the literature. In Switzerland, 41% of people aged 70 years and older use the internet, whereas 97% of people younger than 30 years of age use it [1]. Internet use on a frequent basis (several times a week) is common for 34% of people older than 65 years [2]; the rate in this age group has more than doubled within the last year and is increasing [2,25]. The United States is often seen as a pioneer regarding modern technologies. Already one-half of its population over 65 years uses the internet [3], which can be seen as a prediction for Europe. According to the Nielsen Norman Group [26], approximately 65% of people in the United Kingdom aged 65 to 74 years are using the internet. Older persons in need of support in Europe use the internet primarily for seeking information. Bilateral communication and social interaction via blogs, social networks, and web communities are not yet common among older persons in need of support living in Europe. Only about 10% of older persons in need of support use web 2.0 tools [27]. However, in the United States, 34% of people older than 65 years use social networks for communicating with family and friends [3]. Therefore, it is not surprising that participants in this study are not yet familiar with the potential of ICT. According to the literature, web accessibility currently faces three key challenges: (1) the web is growing faster than accessibility efforts progress; (2) as content, presentation, and design of websites are getting more sophisticated, so must technical skills; and (3) the rise of user-friendly (social) web platforms enabled non–tech-savvy people to share huge amounts of data online and most of them are in inaccessible formats [28]. These aspects should be considered when developing a web platform for older persons in need of support.

Hence, FG participants revealed many challenges when using this web platform. The main obstacles were ethical issues, such as mistrust in data protection and privacy concerns; financial issues, such as costs to purchase hardware and for internet connection; cultural issues, such as age and the knowledge gap for ICT usage; and common issues, such as the assumed information overload on the internet. These findings can be confirmed by the literature [29-31]. The study revealed frustration by users, which can result from searching for information—for example, when a Google search yields too many results. Furthermore, mistrust in the search results was a problem for users. They did not know how to select relevant information. This means that resources and knowledge of quality standards are currently lacking. These challenges of older persons in need of support regarding internet searching are congruent with the literature [8].

One additional factor was named in this study: participants identified social isolation to be a risk for older persons in need of support, which would rise with increased ICT usage. This aspect was mentioned in the context of cooperatives and challenges of using the web platform or ICT together. Furthermore, participants were afraid of older persons in need of support using apps to book a service or to talk to care providers via Skype instead of through a personal, face-to-face conversation. They confirmed that CBHC organizations or care providers in some cases are the last remaining contact that care receivers or older persons in need of support have. Therefore, they ranked personal contact in a CBHC organization as the most important component. They reported that booking personal contact as a service feature was considered taboo. In the literature, there is no agreement on whether social isolation and loneliness increases or decreases as a result of internet use [32]. Various research results indicate that internet usage is connected to a reduction of social isolation among older persons in need of support and increased well-being [33-36]. A decrease in loneliness and increase in social contacts were recently confirmed as benefits of internet usage for people living in assisted living communities [32]. These studies were mainly conducted in the United States. The low rate of internet usage behavior among older persons in need of support in Europe could result in a lack of unawareness of the possibilities related to ICT usage. Intensifying training classes for older persons in need of support can promote ICT usage, especially communication via web 2.0 tools. This could lead to a decrease in social isolation among older persons in need of support. It was especially important for the project team to obtain input from potential web platform users. We were reassured that CBHC services focus on individual contact. The anxiety of older persons in need of support that social isolation may rise with internet usage has to be considered. Recent research shows opposite outcomes, which reveal improved well-being and a decrease in social isolation [32]. Therefore, fostering an increase in ICT usage in older persons in need of support and supporting their growing ICT skills needs be an aim of today's society.

Challenges for CBHC Organizations

In addition to issues related to ICT, stakeholders also identified more general aspects regarding challenges for CBHC organizations. Organizations involved in the study offer a wide

range of services, from accompanying someone to the doctor to mowing their lawn or cleaning their house. One of their principles seems to be to find solutions for any requests from their members or older persons in need of support. In both countries, Switzerland and Slovenia, only informal care was delivered by CBHC organizations. The reason might be the existing structures in the health care system, where professional care is delivered by special outpatient nursing services, such as Spitex in Switzerland, which is paid by insurance companies, the state, and care receivers [15]. The stakeholder group of significant others of care receivers emphasized that one inclusive organization, which offers and coordinates care services of care receivers, would be preferable.

One further issue is the legal form of CBHC organizations. Participants of this Swiss and Slovenian study were predominantly CBHC organizations acting as associations because only a few genuine care "cooperatives" exist in these two countries [37,38]. The International Cooperative Alliance defines a cooperative as an "autonomous association of persons united voluntarily to meet their common economic, social, and cultural needs and aspirations through a jointly-owned and democratically controlled enterprise" [39]. Cooperatives as an organizational form are less recognized in the health care sector in Europe, despite some examples from Italy [40-42]. In Switzerland and Slovenia, cooperatives in the health care sector are still uncommon. Instead, associations act like cooperatives. In Switzerland, for instance, an adaptation of the law would be necessary to enable organizations to found cooperatives and facilitate their work [10]. In Slovenia, the cooperative law was adapted several times, most recently in 2009, when the European cooperative legal order was introduced to Slovenian law [11]. This could be a potential explanation for the small number of existing care cooperatives in Slovenia.

Limitations

While many helpful requirements and needs were identified to develop the platform, some limiting aspects of the study have to be considered. One aspect concerns the study design. The overall framework of the project restricted the number of FGs and involved countries. Nevertheless, the heterogeneity of the countries was satisfying because Switzerland is, in contrast to Slovenia, one of the leading countries concerning technology in Europe. The majority of FG participants being women reflects the societal phenomenon that women are still more involved in caregiving, professionally and informally [43]. Therefore, the FG sampling is assumed to be comparable with the status quo of CBHC organizations. Based on these framing project conditions, the criteria of saturation could not be applied for data collection. Thus, the question cannot be answered whether additional FGs from the same target groups and countries would have yielded more or different information. In the end, one could assert that the FGs revealed enough information because we did not notice any additional information gain from Slovenia compared with Switzerland.

Participants mainly agreed on the proposed workflow model, which was the basis for the FG discussions. Nevertheless, this abstract workflow was a complex model that was not easy to understand for the participants. To compile the lessons learned, more time is required to develop an in-depth understanding and discussion of outcomes. Therefore, we focused the discussions on the questions of the topic guide.

Conclusions

The study revealed a complex variety of results; some of the major issues are summarized below:

- Workflows of CBHC organizations need to be simple and congruent with the discussed workflow model (Figure 1):
  - older persons in need of support and their significant others need information presented in a user-friendly way about the CBHC organization and their services;
  - at best, relevant information about the needs of older persons in need of support from a specific region needs to be presented on the CBHC web platform; and
  - older persons in need of support and care providers need applications to establish contact with each other. Potential tools can vary from telephone calls and personal meetings to online video calls or chat applications.
- Although an overall openness regarding ICT was identified, ICT usage behavior varied enormously among FG participants. This makes it even harder to develop customized web platforms for this target group.
- Challenges of the potential web platform seem to be related to general concerns, such as data protection, access to the internet and ICT, and knowledge gaps in using ICT in the target group.
- The organizational form of CBHC organizations as associations is preferable because of low-threshold structure and establishment. The form of cooperatives is not common in the countries involved in this study.
- Rationales of CBHC organizations are quite clear: maintain a user focus, respect autonomy of older persons in need of support, and maintain personal contact between CBHC organizations and older persons in need of support.

In summary, we conclude that the majority of current stakeholders (older persons in need of support, significant others, and managers/care providers) are not familiar or experienced with ICT usage. Coming generations will contribute more concrete ideas of useful requirements and will benefit more from the web platform. For this reason, training for using the web platform and for setting it up for the specific CBHC organization will be offered. Older persons in need of support seemed very open-minded regarding training for ICT usage; therefore, existing offers of training should be intensified to meet the demands of the older persons in need of support. Once the target users are trained to use ICT, they can experience benefits from the web platform, such as decreased social isolation, independent living, and adequate user-oriented health care services. Further research is required in this context.

The web platform has the potential to facilitate the foundation, work, and collaboration of CBHC organizations in Europe. Overall, personal contact is a main request of older persons in need of support and managers and care providers; in Slovenia, it was cited by almost all participants. Searching for care services, contacting care providers, and communicating with

care providers is preferred via personal contact and seems to be the key element for user acceptance and for the successful implementation of a web platform like “ICareCoops” [16] to support CBHC organizations.

Acknowledgments

We want to thank all participants for their support and contribution to this study. Without their voluntary participation and their openness to discuss their personal experiences and opinions, it would not have been possible to achieve these results. We thank our colleagues David Stamm and Matthew Kerry-Krause for proofreading the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional file requirements matrix.
[XLSX File (Microsoft Excel File), 25 KB - jmir_v23i3e24006_app1.xlsx]

References

1. Bundesamt für Statistik. Erhebung Internetnutzung 2017: Digitale Kompetenzen, Schutz der Privatsphäre und Online-Bildung: die Schweiz im internationalen Vergleich. Neuchàtel: Bundesamt für Statistik; 2018.
2. Seifert A, Schelling H. Digitale Senioren: Nutzung von Informations- und Kommunikationstechnologien (IKT) durch Menschen ab 65 Jahren in der Schweiz im Jahr. Bern: Pro Senectute Schweiz; 2015.
3. Zickuhr K, Madden M. Older adults and internet use. Pew Research Center Internet & Technology. 2012. URL: https://www.pewresearch.org/internet/2012/06/06/older-adults-and-internet-use/ [accessed 2021-02-28]
4. World Health Organization. Global strategy and action plan on ageing and health. Geneva: World Health Organization; 2017.
5. Berlinger N, Solomon MZ. Becoming Good Citizens of Aging Societies. Hastings Cent Rep 2018 Sep 11;48 Suppl 3:S2-S9. [doi: 10.1002/hast.905] [Medline: 30311227]
6. Herzlinger R. Why innovation in health care is so hard. Harv Bus Rev 2006 May;84(5):58-66, 156. [Medline: 16649698]
7. Lupton D. The digitally engaged patient: Self-monitoring and self-care in the digital health era. Soc Theory Health 2013 Jun 19;11(3):256-270. [doi: 10.1057/sth.2013.10]
8. Luzsa R, Williger B, Lang F. Open Innovation mit Senioren. In: Gestaltungsempfehlungen für eine Open Innovation Community. Nürnberg: Fraunhofer-Arbeitsgruppe für Supply Chain Services SCS; 2014.
9. Lam M, McDougle L. Community Variation in the Financial Health of Nonprofit Human Service Organizations: An Examination of Organizational and Contextual Effects. Nonprofit and Voluntary Sector Quarterly 2015 Jun 25;45(3):500-525. [doi: 10.1177/0899764015591365]
10. Helmig B, Gmür M, Bärlocher C, Schnurbein VG, Degen B, Nollert M, et al. The Swiss Civil Society Sector in a Comparative Perspective. Fribourg: The Institute for Research on Management of Associations, Foundations and Cooperatives; 2011:39.
11. Črnak-Meglič A, Rakar T. The role of the third sector in the Slovenian welfare system. Teorija in praksa 2009 Mar 01;46:237-254 [FREE Full text]
12. Brownstein JN, Hirsch GR, Rosenthal EL, Rush CH. Community health workers "101" for primary care providers and other stakeholders in health care systems. J Ambul Care Manage 2011;34(3):210-220. [doi: 10.1097/JAC.0b013e31821c645d] [Medline: 21673520]
13. Colombo F, et al. Help Wanted? Providing and Paying for Long-Term Care. Paris: OECD Health Policy Studies, OECD Publishing; 2011.
14. Creighton H. Europe’s Ageing Demography. In: ILC-UK 2014 EU Factpack. London: ILC–UK; Nov 2014:1-20.
15. Wie ist die NPO-Spitex organisiert? Spitex Schweiz. 2019. URL: https://www.spitex.ch/NPO-Spitex/Struktur-Finanzierung/PAL3U/ [accessed 2021-02-28]
16. iCareCoops. URL: http://project.icarecoops.eu/ [accessed 2021-02-28]
17. Vredenberg K, Isensee S, Righi C. User-Centered Design: An Integrated Approach with CDROM. Upper Saddle River, NJ: Prentice Hall PTR; 2001.
18. Bernhard G, Mahler C, Seidling HM, Stützle M, Ose D, Baudendistel I, et al. Developing a Shared Patient-Centered, Web-Based Medication Platform for Type 2 Diabetes Patients and Their Health Care Providers: Qualitative Study on User Requirements. J Med Internet Res 2018 Mar 27;20(3):e105 [FREE Full text] [doi: 10.2196/jmir.8666] [Medline: 29588269]
19. Krueger R, Casey M. Focus groups: A practical guide for applied research. Thousand Oaks, California: SAGE Publications; 2009.
20. Flick U. Designing Qualitative Research. London: Sage Publications Inc; 2018:20-21.

21. Humanforschungsgesetz. Bundesgesetz über die Forschung am Menschen. Die Bundesversammlung der Schweizerischen Eidgenossenschaft. 2014. URL: https://www.admin.ch/opc/de/classified-compilation/20061313/index.html [accessed 2021-02-28]
22. Ethikkommission in Österreich. Haas J. 2014. URL: http://www.medunigraz.at/ethikkommission/index.html [accessed 2021-02-28]
23. Cohn M. User stories applied: For agile software development. Boston: Addison-Wesley; 2004.
24. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
25. van Eimeren B, Frees B. Drei von vier Deutschen im Netz – ein Ende des digitalen Grabens in Sicht? Media Perspektiven 2011;(7–8):334-349 [FREE Full text]
26. Kane L. Usability for Seniors: Challenges and Changes. Nielsen Norman Group. 2019. URL: https://www.nngroup.com/articles/usability-for-senior-citizens/ [accessed 2021-02-28]
27. Busemann K, Gscheidle C. Web 2.0: Aktive Mitwirkung verbleibt auf niedrigem Niveau. Media Perpektiven 2011;(7–8):360-369 [FREE Full text]
28. Fernandes N, Costa D, Duarte C, Carriço L. Evaluating the Accessibility of Web Applications. Procedia Comput Sci 2012;14:28-35. [doi: 10.1016/j.procs.2012.10.004]
29. Lee B, Chen Y, Hewitt L. Age differences in constraints encountered by seniors in their use of computers and the internet. Comput Hum Behav 2011 May;27(3):1231-1237. [doi: 10.1016/j.chb.2011.01.003]
30. Gatto SL, Tak SH. Computer, Internet, and E-mail Use Among Older Adults: Benefits and Barriers. Educ Gerontol 2008 Aug 22;34(9):800-811. [doi: 10.1080/03601270802243697]
31. Raufi B, Ferati M, Zenuni X, Ajdari J, Ismaili F. Methods and Techniques of Adaptive Web Accessibility for the Blind and Visually Impaired. Procedia - Social and Behavioral Sciences 2015 Jul;195:1999-2007. [doi: 10.1016/j.sbspro.2015.06.214]
32. Cotten SR, Anderson WA, McCullough BM. Impact of internet use on loneliness and contact with others among older adults: cross-sectional analysis. J Med Internet Res 2013 Feb 28;15(2):e39 [FREE Full text] [doi: 10.2196/jmir.2306] [Medline: 23448864]
33. White H, McConnell E, Clipp E, Bynum L, Teague C, Navas L, et al. Surfing the Net in Later Life: A Review of the Literature and Pilot Study of Computer Use and Quality of Life. J Appl Gerontol 2016 Jun 29;18(3):358-378. [doi: 10.1177/073346489901800306]
34. Sum S, Mathews RM, Hughes I, Campbell A. Internet use and loneliness in older adults. Cyberpsychol Behav 2008 Apr;11(2):208-211. [doi: 10.1089/cpb.2007.0010] [Medline: 18422415]
35. Erickson J, Johnson GM. Internet use and psychological wellness during late adulthood. Can J Aging 2011 Jun;30(2):197-209. [doi: 10.1017/S0714980811000109] [Medline: 24650669]
36. Shapira N, Barak A, Gal I. Promoting older adults' well-being through Internet training and use. Aging Ment Health 2007 Sep;11(5):477-484. [doi: 10.1080/13607860601086546] [Medline: 17882585]
37. Quintana CC. The power of cooperation: Cooperatives Europe key figures 2015. Cooperatives Europe. 2016. URL: https://tinyurl.com/84637bfr [accessed 2021-02-28]
38. Girard J. Better Health & Social Care: How are co-ops & mutuals boosting innovation & access worldwide?. Montréal: LPS Productions; 2014.
39. International Co-operative Alliance | ICA. ICA. 2019. URL: https://www.ica.coop/en [accessed 2021-02-28]
40. Mori PA. Community and cooperation: The evolution of cooperatives towards new models of citizens' democratic participation in public services provision. Ann Public Coop Econ 2014 Aug 08;85(3):327-352. [doi: 10.1111/apce.12045]
41. Carini C, Costa E, Carpita M, Andreaus M. The Italian Social Cooperatives in 2008: A Portrait Using Descriptive and Principal Component Analysis. SSRN Journal 2012;35(12):2-23 [FREE Full text] [doi: 10.2139/ssrn.2062407]
42. Thomas A. The Rise of Social Cooperatives in Italy. VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations 2004 Sep;15(3):243-263. [doi: 10.1023/b:volu.0000046280.06580.d8]
43. Boniol M, McIsaac M, Xu L, Wuliji T, Diallo K, Campbell J. Gender equity in the health workforce: Analysis of 104 countries. Geneva: World Health Organization; 2019.

Abbreviations

AAL: Active and Assisted Living
CBHC: community-based health care
COREQ: Consolidated Criteria for Reporting Qualitative Research
FG: focus group
ICT: information and computer technology

Edited by G Eysenbach; submitted 01.09.20; peer-reviewed by T Stamm, R De Carvalho; comments to author 20.09.20; revised version received 01.11.20; accepted 15.01.21; published 10.03.21.

Please cite as:
Biehl V, Becker H, Ogrin A, Reissner A, Burger J, Glaessel A
User-Centered Development of a Web Platform Supporting Community-Based Health Care Organizations for Older Persons in Need of Support: Qualitative Focus Group Study
J Med Internet Res 2021;23(3):e24006
URL: https://www.jmir.org/2021/3/e24006
doi:10.2196/24006
PMID:33688837

Original Paper

A Clinical Communication Tool (Loop) for Team-Based Care in Pediatric and Adult Care Settings: Hybrid Mixed Methods Implementation Study

Amna Husain[1,2,3], MD, MPH; Eyal Cohen[4], MSc, MD; Raluca Dubrowski[5], PhD; Trevor Jamieson[6], MD, MBI; Allison Miyoshi Kurahashi[1], MEd; Bhadra Lokuge[1], MHS; Adam Rapoport[7,8], MD, MHSc; Stephanie Saunders[1,9], MA; Elaine Stasiulis[10], BSc, MA; Jennifer Stinson[5], RN-EC, PhD; Saranjah Subramaniam[11], MA; Pete Wegier[11], PhD; Melanie Barwick[5,12,13], PhD, CPsych

[1]Temmy Latner Centre for Palliative Care, Sinai Health, Toronto, ON, Canada
[2]Lunenfeld-Tanenbaum Research Institute, Toronto, ON, Canada
[3]Division of Palliative Care, Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada
[4]Pediatric Medicine and Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, ON, Canada
[5]Child Health Evaluative Sciences, Research Institute, The Hospital for Sick Children, Toronto, ON, Canada
[6]Department of Medicine, Unity Health Toronto, Toronto, ON, Canada
[7]Pediatric Advanced Care Team, The Hospital for Sick Children, Toronto, ON, Canada
[8]Emily's House Children's Hospice, Toronto, ON, Canada
[9]School of Rehabilitation Sciences, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada
[10]Institute of Medical Science, University of Toronto, Toronto, ON, Canada
[11]Humber River Hospital, Toronto, ON, Canada
[12]Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, ON, Canada
[13]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Amna Husain, MD, MPH
Temmy Latner Centre for Palliative Care
Sinai Health
60 Murray St, 4th Floor
Toronto, ON, M5T 3L9
Canada
Phone: 1 416 586 4800 ext 7884
Email: amna.husain@utoronto.ca

Abstract

Background: Communication within the circle of care is central to coordinated, safe, and effective care; yet patients, caregivers, and health care providers often experience poor communication and fragmented care. Through a sequential program of research, the Loop Research Collaborative developed a web-based, asynchronous clinical communication system for team-based care. Loop assembles the circle of care centered on a patient, in private networking spaces called Patient Loops. The patient, their caregiver, or both are part of the Patient Loop. The communication is threaded, it can be filtered and sorted in multiple ways, it is securely stored, and can be exported for upload to a medical record.

Objective: The objective of this study was to implement and evaluate Loop. The study reporting adheres to the Standards for Reporting Implementation Research.

Methods: The study was a hybrid type II mixed methods design to simultaneously evaluate Loop's clinical and implementation effectiveness, and implementation barriers and facilitators in 6 health care sites. Data included monthly user check-in interviews and bimonthly surveys to capture patient or caregiver experience of continuity of care, in-depth interviews to explore barriers and facilitators based on the Consolidated Framework for Implementation Research (CFIR), and Loop usage extracted directly from the Loop system.

Results: We recruited 25 initiating health care providers across 6 sites who then identified patients or caregivers for recruitment. Of 147 patient or caregiver participants who were assessed and met screening criteria, 57 consented and 52 were enrolled on Loop, creating 52 Patient Loops. Across all Patient Loops, 96 additional health care providers consented to join the Loop teams. Loop usage was followed for up to 8 months. The median number of messages exchanged per team was 1 (range 0-28). The monthly check-in and CFIR interviews showed that although participants acknowledged that Loop could potentially fill a gap, existing modes of communication, workflows, incentives, and the lack of integration with the hospital electronic medical records and patient portals were barriers to its adoption. While participants acknowledged Loop's potential value for engaging the patient and caregiver, and for improving communication within the patient's circle of care, Loop's relative advantage was not realized during the study and there was insufficient tension for change. Missing data limited the analysis of continuity of care.

Conclusions: Fundamental structural and implementation challenges persist toward realizing Loop's potential as a shared system of asynchronous communication. Barriers include health information system integration; system, organizational, and individual tension for change; and a fee structure for health care provider compensation for asynchronous communication.

(J Med Internet Res 2021;23(3):e25505) doi:10.2196/25505

KEYWORDS

coordination of care; complexity; internet communication technology; collaborative care; implementation science; theory of behavior; interprofessional team; patient engagement; social networking technology; user-centered design; Consolidated Framework for Implementation Research; Quality Improvement Framework; Implementation Outcome Taxonomy

Introduction

Background

Collaboration is fundamental to the care of patients with complex needs [1]. Optimal patient outcomes require integrated cross-disciplinary expertise alongside patient and caregiver engagement [2,3]. Effective communication within the circle of care is essential for coordination, cooperation, collaboration, safety, quality, and cost-effectiveness; yet poor communication and fragmented care is too often the norm [2,3].

In this paper, we use the concepts of communication, coordination, cooperation, and collaboration as defined by Fuks et al [4] and as employed in Eikey et al's [5] review of health information technologies and collaboration. Communication is the "exchange of messages and information among people; coordination is the management of people, their activities and resources; cooperation is the production taking place on a shared workspace." Collaboration encompasses communication, coordination, and cooperation, but is much more than its parts [4,5]. Collaboration includes "the development and testing of rules of engagement and shared understanding that facilitates how people work together" [4,5].

A key element of quality care is that patients and families experience good continuity of care (COC); care that is connected and coherent over time [6,7]. Because informational, management, and relational continuity are aspects of COC, effective communication among team members is likely to improve the patient-level outcome of COC.

To promote effective communication, and to examine how this relates to other aspects of collaboration and COC, we developed Loop, a web-based, asynchronous clinical communication tool formerly called My Team of Care (myTOC) [8]. Loop enables private threaded communication among patients, their caregivers, and their health care providers. In this paper, "teams" refers to the members of these digital circles of care called "Patient Loops."

Loop emerged from the need to facilitate open communication between all members of the circle of care, regardless of their role, organizational affiliation, or geographic location. Loop was developed using user-centered design principles. The interface is intuitive [9], allowing patients and caregivers to communicate with their various health care providers (HCPs) in the Patient Loop in a flexible and timely way [10].

To date, although similar tools have been developed and taken to market [11-14], no such tool has been successfully implemented at scale. This raises the question of whether communication tools such as Loop are useful and, if so, what factors might impact their implementation and clinical effectiveness and ultimate scalability.

A pilot randomized controlled trial (RCT) of Loop in patients with advanced cancer demonstrated that Loop was intuitive and usable by members of the patient care team and used as intended for team-based communication in some Patient Loops. There was a nonsignificant trend in improved patient self-reported COC in the intervention group over the 3-month study period [8]. Adoption was influenced by a complex set of system, organizational, team, and individual factors, which is consistent with evidence on determining factors associated with effective implementation [15]. This study examines implementation barriers or facilitators, while also exploring Loop's clinical and implementation effectiveness.

Implementation of eHealth Technologies

eHealth is defined as the application of information, computer, or communication technology to some aspects of health or health care [16]. The widespread use and integration of eHealth interventions into routine care remains a challenge, and most eHealth technologies linger within the confines of the academic settings in which they are studied and are not sustained in practice [17]. Implementation science can address this problem by studying contextual factors [18], process [19], and intervention effects that result in eHealth technologies that are more externally valid, practical, and sustainable, while identifying issues that are important to stakeholders and users.

A theory-informed approach to studying eHealth technology implementation addresses weaknesses reported in existing studies, namely, that they are often based on one particular technology, setting, or health condition, making it difficult to access the available evidence that can inform implementation planning [16]. A recent systematic review of 37 eHealth technologies analyzed using the Consolidated Framework for Implementation Research (CFIR) [15] as an organizing framework recommended that eHealth technology implementation should consider the following highly salient factors: complexity, adaptability, compatibility, cost, and champions. Identifying and monitoring of these barriers can support implementation planning, inform the use of mitigating implementation strategies, and improve implementation effectiveness [20].

Implementation Frameworks

As there is no implementation science model that specifically addresses eHealth technology, we utilized well-established models of implementation to inform the process, factors, and outcomes for this study. The Quality Improvement Framework (QIF) [21] was derived from 25 implementation process frameworks to foster high-quality implementation. The QIF lays out 4 phases that serve as a useful blueprint for the implementation process: phase 1, initial considerations for readiness in the host setting; phase 2, creating a structure for implementation; phase 3, offering the intervention and monitoring ongoing structure; phase 4, sustaining the practice and improving future applications.

The CFIR [15] is a determinant framework comprising 39 key factors associated with successful implementation, structured within 5 domains: intervention characteristics, inner setting, outer setting, characteristics of individuals, and the implementation process. Recent research by Barwick and others [22,23] has identified a subset of factors found to be more salient across contexts. This knowledge can streamline the assessment of barriers toward more effective implementation.

Implementation outcomes [24] are distinct from clinical outcomes and capture effects of deliberate actions to implement interventions in new settings. Implementation outcomes have 3 important functions: (1) they serve as indicators of implementation success; (2) are proximal indicators of implementation processes; and (3) are key intermediate outcomes in relation to clinical outcomes. Implementation outcomes include acceptability, adoption, appropriateness, cost, feasibility, fidelity, penetration, and sustainability. When interventions fail to produce desired outcomes, it is important to know if the failure occurred because the intervention was ineffective (intervention failure) or whether it was implemented incorrectly (implementation failure).

Objectives

The study examined the implementation and clinical effectiveness of Loop across 6 health care settings. We assessed clinical outcomes (COC, client participation in decision-making), implementation outcomes (adoption, acceptability, appropriateness, and feasibility), and explored implementation barriers and facilitators. We hypothesized that an implementation approach informed by the core principles of implementation science (ie, process, factors, strategies, outcomes, and implementation team) would lead to adoption, and that higher Loop use would be associated with improved patient COC. We anticipated identifying similar salient determinant factors that have been documented across varied study contexts and interventions. The study reporting adheres to the Standards for Reporting Implementation Research (Multimedia Appendix 1) [25].

Methods

Study Design

The study design was a hybrid type II, involving the simultaneous testing of a clinical intervention and an implementation strategy with the aim of more rapid translation [26]. We used a mixed methods approach to examine the clinical and implementation effectiveness of Loop at 6 health care sites [26]. The study was approved by the Research Ethics Board of Sinai Health System, University Health Network, and SickKids Hospital, and was conducted in Toronto, Ontario, Canada, where health care is provincially funded.

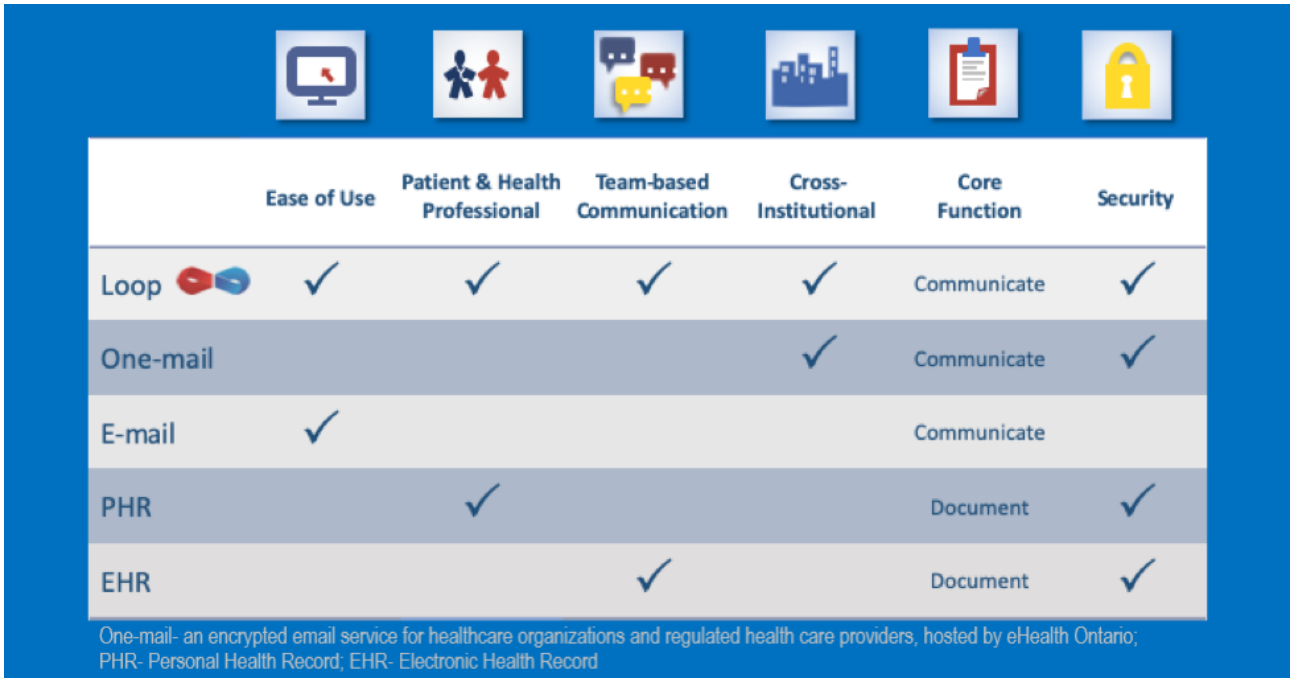
Loop Intervention

Loop enables private communication groups centered on a patient, called Patient Loops (Figures 1 and 2). In each Patient Loop, there are 2 streams of communication, one that includes the patient and caregiver, and another that is for the health care providers only [27]. Messages are threaded for ease of viewing conversations. Messages may be tagged with customizable labels (eg, hypertension, pain, lymphedema), and marked to the attention of a specific member or members of the Patient Loop. The latter action triggers a deidentified link to be sent to the email of the intended recipient(s) [28]. Figures 1 and 2 depict the Loop interface, and Figure 3 shows how Loop functionality compares to other categories of eHealth tools.

Figure 1. Screenshots of Loop optimized for a smartphone.

Figure 2. Screenshot of Loop on a computer screen.

Figure 3. Comparison of Loop with other categories of eHealth tools.



Site Recruitment

Six clinical sites participated in this study. All sites were in academic health organizations in Toronto, Ontario, Canada. Three clinical sites were recruited in the first roll-out, including a regional palliative care program (Site 1) that provides home-based palliative care, alongside home care organizations; an academic family health program (Site 2) that provides primary care to patients; and a brain metastases clinic (Site 3) housed within an outpatient regional cancer center. These first 3 sites were approached during the knowledge translation activities for the previous stages of the research program. For this study, we reached out to health care colleagues in various specialties to present the study aims to site leads. Once site leads expressed interest in participating, we presented to the broader clinical group at each site.

In a second roll-out, we recruited a pediatric blood and marrow transplant (Site 4) program, and a pediatric palliative care program (Site 5)—both pediatric sites are situated within the same quaternary pediatric teaching hospital; and an outpatient psychosocial oncology program (Site 6) at a regional cancer center. The sites recruited in this second roll-out approached us, having learned of the study from colleagues or the Loop Research Collaborative. Five of the sites were specialized in hematology-oncology, radiation-surgical oncology, psychosocial oncology, or palliative care. The family medicine program was included to examine Loop adoption in primary care. An implementation champion was identified at each site. All champions were clinicians, most had a leadership role, and they engaged other HCPs at their site to elicit participation in the study.

Recruitment of HCPs, Patients, and Caregivers

The implementation champion at each site identified initiating HCPs (iHCPs) for recruitment. Additionally, study staff identified and recruited iHCPs at implementation planning activities described below. iHCPs then identified patients or their caregivers who were screened for inclusion criteria. Once the iHCP and the patient were registered in Loop, both were asked to identify any additional HCPs from the patient's circle of care who could be invited to participate on the Loop. Study staff or iHCPs invited additional HCPs via email, phone, and in-person. There was no limit to the number of additional HCPs invited to join the patient's Loop, and all provided verbal consent upon joining. Study staff followed a standard procedure to invite and enroll participants. The Loop Help menu contains videos for a Loop "quick start" guide.

Inclusion Criteria

Patients/caregivers from adult centers were included if (1) they were aged 18 or older and had capacity to consent. Pediatric patients (18 years or younger) could consent themselves, if capable, otherwise their parent or guardian was consented; (2) the patient or caregiver had internet access; and (3) there were at least two HCPs involved in the patient's care. An additional criterion for patients recruited from adult centers was an Eastern Cooperative Oncology Group (ECOG) Performance Status score of 2 or less [29,30]. There was no comparable performance status measure for pediatric patients.

Exclusion Criteria

Patients were excluded if they (1) they had a prognosis of less than 6 months as determined by a physician, except for adult and pediatric palliative care sites where it was difficult for clinicians to identify those that met this criterion; or (2) had cognitive impairment as determined by a physician or by study staff using the Bedside Confusion Scale for adult patients [31].

Sample Size

No sample size calculation for clinical effectiveness was possible due to the limited sample size of the feasibility trial. Based on the previous study in ambulatory cancer and palliative care [8], we anticipated it would be feasible to recruit 15 teams

during the first wave of recruitment at 3 sites and 5 teams from each additional site during the second wave, resulting in 60 teams or Patient Loops. Based on a 25% attrition rate at various steps of enrollment, onboarding, and assembling the team, we anticipated a total enrollment of 45 Patient Loops in this study.

Implementation Procedure

The initiating context for this implementation endeavor was research. The intention was to provide Loop to participating organizations with the aim of exploring implementation and clinical effectiveness. We did not set out to implement Loop within entire organizations. As such, we did not undertake certain implementation activities such as developing organizational implementation teams and ensuring sustainability, as these are process elements key to program- or organizational-level implementation. Previous research [32] has identified that initiating context or impetus for the implementation endeavor is important for implementation process and sustainability.

Implementation Phase 1

Phase 1 occurred over 3-6 months and focused on understanding the initial implementation considerations within each site (QIF Phase 1) [21]. An assessment of needs, capacity, and readiness was done at each site, led by study staff, and guided by the Hexagon Tool [33]. The purpose of these meetings was to explore process adaptations that might be required, clarify goals, provide information about collaborative care and Loop, and to establish buy-in for using Loop. We conducted workflow observations to understand HCPs' clinical workflows. Focus groups and consultative meetings were held to introduce and refine a tailored implementation plan for the study. HCPs at each initiating site were invited to participate in the information meetings where they were recruited as iHCPs, registered on Loop, and baseline data were collected.

Implementation Phase 2

Phase 2 spanned 3-6 months and focused on infrastructure and workflow adjustments for implementing Loop (QIF Phase 2) [21]. Using the description drawn from the implementation framework, site implementation champions were identified among HCPs who expressed interest in this role, although no formal role designation was made at the program level. Site readiness assessments from phase 1 informed a general implementation plan for each site, which was discussed with site champions and HCPs for refinement. iHCPs at all sites were asked to identify patients who met inclusion criteria. Patients or, if appropriate, their caregivers were then consented, enrolled in the study, and registered on Loop, creating a Patient Loop. Patient participants could identify a caregiver to participate in the study, who was also consented and joined the Patient Loop. Patients and their iHCPs were asked to identify other members of the patient's circle of care (additional HCPs) who were invited to join the Patient Loop and were verbally consented. At the time of registration on Loop, study staff showed new Loop users how to use Loop.

Implementation Phase 3

Phase 3 focused on initiating Loop use at the site and providing ongoing supports [21]. Loop was available to each site and team for up to 8 months, during which data were collected. Implementation strategies were used to maintain user engagement, including (1) monthly check-in phone calls with participants to gather information on user experiences, provide support, and troubleshoot Loop use; and (2) periodic audit and feedback summaries on Loop uptake posted by study staff in each Patient Loop at bimonthly intervals. Feedback summaries were intended to remind users to use Loop and as a positive peer pressure stimulus to encourage Loop use.

Implementation Phase 4

For this study, Phase 4 involved an ongoing process of reflection on future applications for Loop [21] and was concurrent with all phases. Study participants, implementers, and stakeholders reflected on implementation process, Loop use, and improvements. These reflections were captured during regular interviews and periodic stakeholder consultations.

Data Collection

Participant Characteristics

Site characteristics were gathered during Phase 1 activities and culminated in an implementation plan for each site. Baseline data for participant characteristics included demographics, internet preferences, performance status (Palliative Performance Scale) [34], and Age-Adjusted Charlson Comorbidity Index (ACCI) [35] for adult patient participants. In a sample of cancer patients, ACCI scores have been categorized as mild (0-1), moderate (2-3), and severe (≥4), corresponding to a significant difference in survival rates [35]. For patients recruited from adult centers, the iHCP determined if patients had high unmet health or social needs as defined by Schaink et al [36]. Although a Pediatric Comorbidity Index is being developed, there is currently no validated measure of comorbidity or complexity for pediatric patients [37].

Implementation Outcomes

Adoption

Adoption is defined as the intention, initial decision, or action to try or employ an intervention or evidence-based practice [24]. In this study, adoption was operationalized as a function of Loop use: (1) the number of patient care teams registered on Loop, (2) the number of participants in each user category on Loop, (3) the total number of messages by site, and (4) the median number of messages per team per site. Loop use metrics were collected from Loop software reporting and backend data export at an interim point and at the end of the study, and by participant self-report in the check-in interviews.

Acceptability

Acceptability is the perception among implementation stakeholders that a given treatment, service, practice, or innovation is agreeable, palatable, or satisfactory [24]. This was assessed informally in the phase 1 preparatory meetings and within the CFIR interviews.

Appropriateness

Appropriateness is the perceived fit, relevance, or compatibility of the intervention for a given practice setting, provider, or consumer; or perceived fit of the innovation to address a

particular issue or problem [24]. This was assessed informally in the phase 1 preparatory meetings and in the CFIR interviews and monthly check-ins.

Feasibility

Feasibility is defined as the extent to which a new intervention can be successfully used or carried out within a given setting [24]. The feasibility of implementing Loop was assessed at the site level with respect to number of sites approached; number of sites who approached us with an interest in participating; number of HCPs interested and recruited; number of HCPs who identified patients for recruitment; number of patients, caregivers, and additional HCPs recruited; and number of active Patient Loops assembled.

Cost, fidelity, penetration, and sustainability were not measured in this study. Fidelity to the intervention, or the extent to which users adhered to the Loop tool as intended [24], did not apply because as an eHealth technology, Loop does not have optional multiple core components; rather, a message is sent or not. Penetration and sustainability [24] were not relevant because Loop was only made available to a discrete number of teams for the purpose of this study, and there was no intention of full implementation within each site as part of usual practice.

Barriers and Facilitators

Barriers and facilitators to Loop implementation were assessed using individual interviews informed by implementation outcomes [24] and CFIR [15] using 2 main qualitative approaches. A brief (5-20 min) monthly check-in phone call with patients, caregivers, and iHCPs was used to gather feedback and troubleshoot implementation and technical issues. Participants were asked about their Loop use over the previous month, and their perception of its acceptability, accessibility, usefulness, feasibility, including the impact of Loop use on workflow, and their willingness to continue using Loop beyond the study if given the opportunity. Study staff conducted and captured monthly check-in content in fieldnotes. Check-in phone calls were audio recorded and reviewed to support fieldnote rigor.

Semistructured interviews based on the CFIR and adapted for language and context were conducted by telephone to capture HCP perspectives on implementation barriers and facilitators (Multimedia Appendix 2). All CFIR domains and constructs were included except for trialability, as this factor did not apply in a research-initiated implementation endeavor. In addition, given the role of patients in the use of Patient Loops, we included a sixth domain related to Patient Beliefs and Experiences to capture HCP perspectives on how patients experienced Loop, which has been done in previous studies [23]. The interview protocol was piloted with 2 HCPs and revised for length, flow, and clarity. Interviews were conducted with the site lead and an additional iHCP from each site who were purposefully sampled to capture sites having higher and lower Loop use. CFIR interviews were 30-60 minutes long, conducted by 2 members of the research team experienced in CFIR interviews and analysis (ES and RD), and supervised by the implementation science lead (MB).

Clinical Outcomes

Clinical outcomes were collected at baseline and at 2-month intervals from all patients or caregivers, either by phone or in person using standardized measures administered via survey (Table 1). Measures assessed patient or caregiver experience of COC, symptom severity, and participation in decision making and goal setting. Internally developed questionnaires measured team effectiveness [38]. Details on circle of care communications occurring outside the Patient Loop were collected from iHCPs, patients, and caregivers at monthly intervals using an internally developed social network questionnaire and will be the focus of a separate paper. Clinical effectiveness outcome measures were not collected from the noninitiating (additional) HCPs to decrease respondent burden and encourage enrollment on Patient Loops.

Table 1. Patient and caregiver outcome measures.

Construct	Survey	Validated in	Scoring details	Administered to
Continuity of care experience (COC)	Continuity and Coordination subscale, Picker Ambulatory Cancer Care Scale [7]	Patients with cancer	Range: 0-100 Higher scores indicate higher continuity of care	Adult and pediatric patients and their caregivers
Symptom severity	Edmonton Symptom Assessment Scale (ESAS) [39,40]	Adult patients with cancer	Range: 0-90 Higher scores indicate higher symptom severity	Adult patients
Symptom bother	Symptom Screening in Pediatrics Tool (SSpedi) [41,42]	Children with cancer and hematopoietic stem cell transplantation	Range: 0-60 Higher scores indicate higher levels of bother	Pediatric patients and their caregivers
Client participation in decision making and goal setting (CPDG)[a]	Client-Centered Rehabilitation Questionnaire (CCRQ) [43], CPDG domain	Discharged rehabilitation patients	Range: 0-100 Higher scores indicate more positive responses	Adult and pediatric patients and their caregivers

[a]CPDG: Client participation in decision-making and goal setting domain of Client Centered Rehabilitation Questionnaire (CCRQ).

Analysis of Participant Characteristics

Participant characteristics (patients, caregivers, and HCPs) are described by site using frequencies, medians or means, SDs, and ranges.

Analysis of Implementation Outcomes

Adoption

Loop adoption was dependent on the number of individuals in each Patient Loop. A Patient Loop was considered active if it included an iHCP and a patient or caregiver. We conducted a subanalysis of the proportion of Loops with at least one additional HCP as part of the care team assembled on Loop.

Monthly Check-in Interviews (Adoption, Acceptability, Appropriateness, Barriers, and Facilitators)

Monthly check-in data were analyzed using hybrid data-theory-driven content analysis [44] on MAXQDA 2018.2. An initial codebook based on monthly check-in questions was developed and iteratively revised throughout the analytic process. The first phase of coding involved 6 members of the project team (AH, MB, PW, StS, SaS, and BL) who independently coded the same 3 monthly check-in interview notes, followed by a discussion to achieve consensus and identify revisions to the codebook. Each coder then rated the same notes from 3 new interviews, which were again reviewed for consensus and codebook revisions. The 6 coders continued coding the remainder of the interview notes independently and met regularly to discuss any new codes and issues related to implementation. Following coding completion, 1 coder (AH) reviewed monthly check-in notes from all sites, identified common coding themes, and summarized excerpts using data-driven content analysis [45]. In addition, some divergent perspectives were analyzed to provide a range of perspectives. Excerpts and summaries were discussed by the 6 coders to achieve agreement on emergent summative statements. The main coder (AH) then synthesized summative statements from text segments within categories and further sorted and analyzed all coded segments by site. A second reviewer (MB) reviewed the summary tables, consisting of coding categories, exemplar excerpts from the interview notes, and summative statements by site.

CFIR Interviews (Acceptability and Appropriateness)

CFIR interviews were analyzed using an adapted rapid analysis method [46]. ES conducted the interview while RD simultaneously coded each interview against CFIR constructs using a pre-set template. After each interview, ES checked her notes against the audio recording for completeness, and then RD sent her coded notes to ES to do a final check for accuracy and completeness. In a second step, not part of the rapid analysis method but consistent with previous CFIR research [47], ES and RD independently assigned valence ratings to each construct based on its strength (–2, –1, 0, 1, 2) and direction (negative or positive) relative to Loop implementation. Disagreements on valence ratings were resolved by discussion and consensus with MB.

Implementation barriers and facilitators were compared within and across sites and CFIR constructs were explored as a function of high and low Loop use. Data overlap between CFIR constructs and monthly check-in data were explored using a mixed methods approach to achieve greater contextual understanding [48].

Analysis of Clinical Effectiveness Outcomes

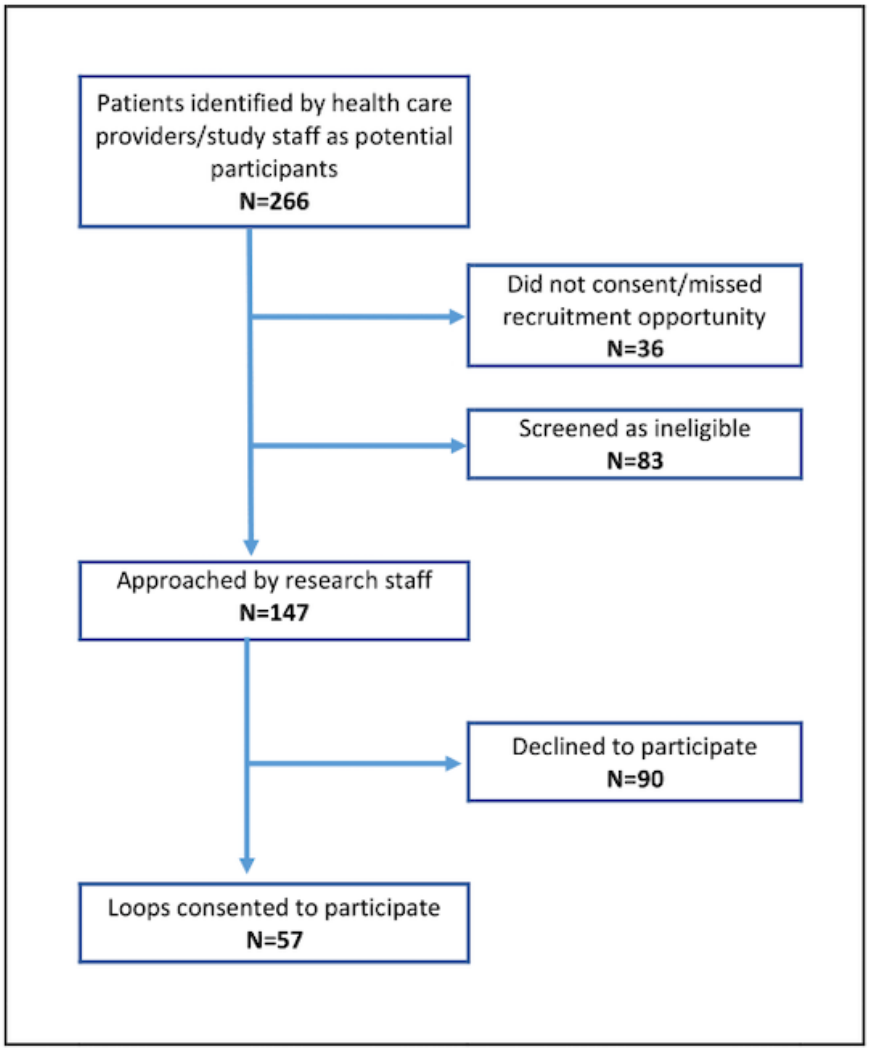
We report descriptive statistics (means, SD, and ranges) for clinical outcome measures. We conducted an analysis of change in score of the main outcome of the Continuity and Coordination Subscale of the Picker (COC) between baseline and each timepoint. We did an exploratory repeated measures analysis of COC controlling for patient participation in decision making (CPDG domain of Client-Centered Rehabilitation Questionnaire [CCRQ]) and Loop use within the Patient's Loop.

Results

Recruitment

A total of 57 HCPs and key informants took part in Phase 1 activities. We recruited 25 iHCPs across all sites who then identified 266 patients as potentially meeting participation criteria. Figure 4 charts the steps in participant recruitment (patients or caregivers) from which Loops were formed.

Figure 4. Patient recruitment flowchart.



Of the 147 patient participants (or caregivers in their lieu) who were assessed and met screening criteria, 57 consented and 55 Loops were created, within which 51 patients participated in data collection. Of the 55 Loops created, attrition resulted in 31 Loops completing the follow-up period. Patients and iHCPs together identified 190 unique additional HCPs who were part of the patient's circle of care (some HCPs were included in more than 1 Patient Loop). Research staff contacted each identified additional HCP an average of 4 times, using phone and email, to invite them to participate in the study. Of these individuals, 96 (50.5%) consented to join a Loop. Of the remaining additional HCPs, 47/190 (24.7%) did not respond to invitations by the research team, 30/190 (15.8%) declined to join, and 17/190 (8.9%) were unable to participate further in the study as the referent patient had died. We did not monitor reasons for declining to join.

Participant Characteristics

Baseline participant characteristics are presented in Tables 2 and 3. Of 51 patients for whom baseline data were collected, 59% (30/51) were female. Patients ranged in age from 1.4 to 90 years. Most patients had a cancer diagnosis (61%, 31/51), although the primary diagnoses ranged from pediatric genetic disorders to connective tissue diseases. For adult patients, performance status was collected with the Palliative Performance Scale, for which median scores ranged from 60% at Site 1 to 80% at Site 6. The minimum Palliative Performance Scale score was 50% and the maximum score was 100% across all sites. The median ACCI comorbidity score ranged from mild in primary care patients to severe among home palliative care patients, demonstrating variable morbidity–mortality within the sample.

Table 2. Patient and caregiver characteristics.

Characteristics	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Overall sample
Participant, n							
Patient	6	10	16	3	5	11	51
Caregiver	2	0	4	1	5	3	15
Gender, n							
Patient, n							
Female	3	7	11	0	2	7	30
Male	3	3	5	3	3	3	20
Other						1	1
Caregiver, n							
Female	1	—	4	—	—	0	5
Male	1	—	0	—	—	3	4
Age (years), median (range)							
Patient	68 (58-90)	62 (18-87)	58 (29-73)	17 (11-27)	14.5 (1.4-17)	52 (25-72)	
Caregiver	73 (67-79)	—	39.5 (38-60)	—	—	65 (54-66)	
Performance status, median % (range)	60 (50-80)	90 (60-100)	80 (60-100)	—	—	80 (70-100)	
Patient diagnoses, n	6	10	16	3	5	11	51
Cancer	6	1	13	0	4	7	31
Noncancer	0	9	3	3	1	4	20
Age-Adjusted Charlson Comorbidity Index (ACCI), median (range)	6 (3-8)	3 (0-9)	5 (0-9)	—	—	2 (0-6)	
Severity (median)	Severe	Mild	Severe	—	—	Mild	
Complexity, n	2	5	15	—	—	9	31
Multimorbidity	2	5	13	—	—	7	27
Resource utilization	2	2	12	—	—	6	22
Psychosocial issues	1	1	2	—	—	9	13

Table 3. Health care provider characteristics.

Characteristics	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Overall sample
Health care provider, n							
Initiating HCP[a] (iHCP)	6	6	4	2	3	4	25
Additional HCP	11	14	26	6	19	20	96
iHCP gender							
Female	2	3	1	1	2	3	12
Male	4	3	3	1	1	1	27
iHCP age (years), median (range)	41 (30-68)	39 (32-67)	40.5 (32-44)	62 (62)	37 (36-39)	49.5 (33-54)	
iHCP type, n							
Physician	6	5	4	2	—	4	21
Clinical nurse specialist (CNS)	—	—	—	—	2	—	2
Nurse practitioner (NP)	—	1	—	—	1	—	2
Clinical specialty	Focused palliative care practice (N=6)	Family physician (N=5); NP (N=1)	Radiation oncology (N=3); Neurosurgeon (N=1)	Pediatric hematologist/oncologist (N=2)	Palliative CNS (N=2); palliative NP (N=1)	Psychiatry (N=4)	
Role, n							
Administrative director	—	1	—	1	—	—	2
Clinical programs director	2	3	1	—	—	1	7
Clinical care	6	6	4	2	3	4	25
Years in health care, median (range)	10.5 (3-37)	12.5 (4-38)	14.5 (6-18)	28 (24-32)	13 (11-15)	15 (6-19)	
Practice fee structure							
Fee for service	—	1	1	—	—	4	6
Alternate payment plan	6	3	—	1	—	—	10
Salaried	1	2	2	1	3	—	9
Other, academic, or alternate funding plan	—	Capitated alternate payment plan (n=3)	Fee for service/alternate funding plan (n=1)	—	—	—	—

[a]HCP: health care provider.

For 4 of 6 sites (Sites 1, 2, 3, and 6), iHCPs were asked to assess patient complexity based on the categories identified by Schaink et al [36]. The majority of patients were identified as having multimorbidity (87%, 27/31) and high resource utilization (71%, 22/31), while a minority were identified as having psychosocial issues (42%, 13/31).

Although internet access was an inclusion criterion, 1 out of 51 patient reported no internet access. Almost all patients and caregivers reported feeling "comfortable" or "very comfortable" using computers, and most felt "comfortable or very comfortable" using a smartphone. Of 51 patients or their caregivers, 24 (47%) used email and 16 (31%) used text to communicate with their HCPs; the remainder communicated by phone, in-person, or by pager.

Implementation Outcomes

Adoption

Loop adoption was based on use statistics pulled from Loop's data server. Across all sites, the total number of participants, including additional HCPs, who joined Patient Loops was 262, with 52 Patient Loops created. The median number of HCPs per Patient Loop was 4 with a range of 1-13 (Table 4). Overall, 228 Loop messages were sent by patients, caregivers, and health care providers within the study period. A full breakdown of

messages sent by user type and site is presented in Table 5. Across sites, a median of 1 message and a maximum of 28 messages were sent within a single team. Quartiles were calculated to characterize the number of messages as low, medium, and high Loop use based on the number of messages exchanged at each site: Q1 (low), Q2 (medium), and Q3 (high). Patients were the most active users, posting nearly 50% of all messages within Loop.

Table 4. Team composition by Patient Loop (team) and by site.

Composition	Site 1[a]	Site 2[b]	Site 3[c]	Site 4[d]	Site 5[e]	Site 6[f]	All sites
Members[g], n	26	27	109	14	49	37	262
Teams, n	6	9	18	3	5	11	52
HCP[h] per team							
Mean (SD)	3.17 (2.14)	2.00 (1.00)	4.89 (0.96)	3.33 (1.15)	7.80 (4.32)	2.09 (1.22)	3.79 (2.44)
Median (range)	2 (1-7)	2 (1-4)	5 (4-7)	4 (2-4)	7 (2-13)	2 (1-4)	4 (1-13)
≥1 secondary HCP	5	5	18	3	5	7	43
Members per team							
Mean (SD)	4.33 (2.16)	3.00 (1.00)	6.06 (1.11)	4.67 (0.58)	9.80 (4.32)	3.36 (1.36)	5.04 (2.62)
Median (range)	3 (2-8)	3 (2-5)	6 (5-8)	5 (4-5)	9 (4-15)	3 (2-5)	5 (2-15)

[a]Six initiating HCPs were recruited from among 18 physicians within an expert palliative care program that has a large homecare component.

[b]Six initiating HCPs were recruited from among 12 physicians in an academic family medicine site.

[c]Three radiation oncologists and 1 neurosurgeon (iHCPs) were recruited from within a multidisciplinary program based in a regional cancer center, which included additionally 2 neurosurgeons, 1 registered nurse (RN), 1 physician assistant (PA), and 1 fellow in training. The PA and Fellow participated as additional HCPs on the Patient Loops.

[d]Two out of 5 physicians, 4 patients, and 2 caregivers were recruited from a Pediatric Blood and Marrow Transplant program within a quaternary pediatric hospital. Additionally, this program has 3 nurse practitioners (NPs) and 4 RNs.

[e]One NP and 2 clinical nurse specialists were recruited as iHCPs from a pediatric palliative care program, which includes 5 physicians, 1 nurse practitioner, and 2 clinical nurse specialists within a quaternary pediatric hospital.

[f]Four out of 9 psychiatrists were recruited. Additionally, this adult psychosocial oncology program located within a regional cancer center has 16 social workers, 5 clinical psychologists, and 2 music/art therapists.

[g]Includes patient, caregiver, and HCP members.

[h]HCP: health care provider.

Table 5. Message frequency by site, user type, and Patient Loops (teams).

Message Frequency	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	All sites
Messages (not including research administrator), N	39	80	62	9	26	12	228
Median (range)	3.5 (0-22)	3 (0-27)	1.5 (0-28)	4 (0-5)	1 (0-18)	0 (0-7)	1 (0-28)
Frequency quartiles (Hi, Med, Lo)	Q3 (Hi)	Q3 (Hi)	Q3 (Hi)	Q1 (Lo)	Q2 (Med)	Q2 (Med)	
Messages sent by user type, n (%)[a]							
Total messages (including research administrator), N	54	98	97	14	30	25	318
Patient	12 (22.2)	53 (54.1)	28 (28.9)	1 (7.1)	0 (0.0)	7 (28.0)	101 (31.8)
Caregiver	18 (33.3)	0 (0.0)	14 (14.4)	3 (21.4)	14 (46.7)	1 (4.0)	50 (15.7)
Health care provider	9 (16.7)	27 (27.6)	20 (20.6)	5 (35.7)	12 (40.0)	4 (16.0)	77 (24.2)
Research admin	15 (27.8)	18 (18.4)	35 (36.1)	5 (35.7)	4 (13.3)	13 (52.00)	90 (28.3)
Teams ≥1 message, n	5	7	10	2	3	5	33

[a]The sum of all messages, including those sent by the Research Admin, were used as the denominator when calculating the % in this section.

Participants also reported their time spent on Loop in the monthly check-in interviews. We had 250 responses to the question of "Loop use in the previous month"; of these, 95 (38.0%) responses reported "some Loop use" over the previous month, and 155(62%) reported "no Loop use" over the previous month. A participant likely responded at more than 1 timepoint to this question, and therefore, these are not independent responses.

Acceptability

Perception of Loop's acceptability [24] as agreeable, palatable, or satisfactory was explored in Phase 1 of implementation, which centered on program composition and workflows, and identification of a site champion who could facilitate buy-in and recruitment. Using an iterative consultative planning process, each site arrived at a decision to proceed or to not proceed with Loop. Among the sites we initially approached, one site did not proceed to Phases 2 and 3 due to competing priorities with the roll-out of another eHealth tool. Among the sites that proceeded to Phases 2 and 3, all participants perceived Loop, the implementation plan, and study procedures as acceptable. Perspectives on acceptability at midpoints and study end are presented in the monthly check-in and CFIR interview analyses below.

Appropriateness

Loop's perceived fit, relevance, or compatibility for setting and gap in care is reported in the monthly check-in and CFIR analyses below.

Feasibility

The extent to which Loop was successfully used within a site was operationalized by recruitment and use statistics, reported in Tables 4 and 5. These data show that Loop was used in each of the participating sites to some extent. It is important to note, however, that sites were recruited within a research context, and site participants may have been motivated to use Loop for this reason.

Barriers and Facilitators

Monthly Check-In Interviews

Monthly check-in interviews captured barriers and facilitators related to Loop use, ways in which Loop filled gaps in care, opportunity to use Loop, team composition and patterns of communication, Loop design and function, and overall satisfaction with Loop. Field notes, reported as excerpted first-person statements below, shed light on loop acceptability and appropriateness.

Loop Use

Among barriers to Loop use, iHCPs, patients, and caregivers reported that existing modes of communication, such as phone, in person, and email were commonly used for medical care needs. As such, participants did not perceive a relative advantage [15] to using Loop for team-based communication, and medical issues were not sufficiently complex to warrant the use of Loop.

> *It's not necessary to add new people on Loop...We have a system that is working well... Typically, just calling the nurse practitioner and/or emailing the nurse at the hospital works. I see the value in the Loop system, but it's not necessary for where I am at.* [Fieldnote, Parent Caregiver, Site 5, Month 2]

Gap in Care

Loop's inclusion of patients in team communication was perceived to be facilitative to Loop use and was identified as fulfilling a gap in care.

> *There is a strong advantage to having the patient in the Loop and being privy to these conversations.* [Fieldnote, iHCP, Site 2, Month 1]

The iHCP quoted above expressed how communication is a challenge, even when programs are part of the same organization and located in the same building. The conventional transfer of information via consult notes does not address this gap. Additionally, iHCPs said that Loop addressed a collaboration gap across the health care team. Patients also expressed that collaboration, specifically, is critical but lacking in the care they experience.

> *The collaborative care element is key, this is what is missing from the patient's experience of the health system…it's a crucial gap that Loop could fill, it's just getting people on board to use it. There needs to be communication between different providers and different sites. The lack of communication leads to care being incredibly fragmented.* [Fieldnote, Patient, Site 6, month 6]

Several participants perceived Loop as having potential to improve their medical care and to prevent the duplication of communication and services. In addition, Loop could provide a means to ask questions or provide updates that may not have been communicated during in-person visits.

> *I found that providers were 'duplicating' some of the same treatment issues and that maybe Loop could be useful for this.* [Fieldnote, Caregiver, Site 1, Month 4]

> *Everyone is on Loop so I am feeling better about not needing to double up with messages. Anyone on the Loop can prescribe if the patient needs something. Any person on the Loop can do the duty needed, which is 11 people on Loop. I find that at home, we have so many services going on that one thing gets mixed up and all of a sudden all the information is wrong and I get stressed out. As a result, I feel like I am not sure what's going on. Loop could help.* [Fieldnote, Caregiver, Site 5, Month 2]

Opportunity to Use Loop

Participants across all sites frequently reported that no medical situation arose during the data collection period that prompted them to post a message in Loop. Patients were either medically stable or in remission and, therefore not requiring active treatment; or they were too sick to use Loop, admitted to hospital or a palliative care unit. Patient and caregiver Loop use was facilitated by instances when a specific situation arose, such as an emergency department visit, a need to coordinate an admission to long-term care, or to ask a question about symptoms. In other instances, a patient or caregiver used Loop to update the health care team, primarily about appointments they had scheduled.

Team Composition and Patterns of Communication

Although some users reported that partial teams could still be useful, assembling additional HCPs in a Patient's Loop proved challenging. Participants frequently stated that unless the relevant team members were enrolled, Loop had limited usefulness.

> *Loop on the other hand is very simple and easy. The goal is to have one place for all the specialists; one place they can go to communicate. I feel that if you can't get everyone to sign on, then it limits the usefulness of Loop.* [Fieldnote, Caregiver, Site 6, month 1]

Participant messages that were left unreciprocated also posed a barrier to Loop use.

> *After trying this and getting no response, I didn't want to use Loop more because I didn't want to feel that I was badgering others.* [Fieldnote, Patient, Site 6, month 6]

Some iHCPs reported that they did not post messages unless patients posted first, perceiving the patients or caregivers as drivers of care-based communication. Of note, in at least one instance, an adolescent patient stated that he would not post messages in Loop unless the HCPs posted first. Patients and caregivers whose messages were reciprocated indicated that they were likely to use Loop again.

Design and Function

Loop's user interface was generally considered to be simple and intuitive, and the asynchronous nature of the messaging useful for nonurgent messages, and thereby, likely to reduce burden. Nonetheless, patients and caregivers who had access to their hospital's patient portal, which allows them to view reports and test results, found the portal met many informational needs, if not their communication needs. Some users expressed that integration of Loop into the patient portal would be helpful.

> *I really, really, like the simplicity of Loop's design, and I feel that it is simple to access for those that might not be tech savvy.* [Fieldnote, Patient, Site 1, month 2]

There was some confusion among participants about Loop's purpose and the types of messages that were appropriate to post.

> *I felt unsure about what concerns can be put on the system. Right now, I generally send emails to my HCPs regarding care plans. There are 12 members in my Loop, but no activity.* [Fieldnote, Caregiver, Site 5, month 2]

iHCPs commonly believed Loop should ideally be integrated into the hospital's electronic medical record (EMR). Because Loop requires its own login and is not embedded in the EMR, its use was cumbersome and did not align with their existing workflow, particularly if only a few of their patients were using it.

Implementation Context

Given that research was the initiating context for the implementation endeavor, research team members played key roles in supporting implementation that would not be sustainable otherwise. Study staff helped users to register on Loop, explaining what tagging "attention to" someone means when posting a message, and clarifying what kinds of messages were

appropriate to post. Study staff posted bimonthly messages with audit information such as the number of messages posted in the participant's Patient Loop, number of messages in the most active Patient Loop across all sites during the same period, how to use Loop, and updates about study.

Workflow and Compatibility

iHCPs had pre-existing processes or workflows that were supported by administrative or other clinical staff. In some settings, clinical administrative staff or trainees were tasked with communicating with patients and other HCPs, which meant HCPs did not experience the back and forth "telephone tag" that is common when communicating with patients and other HCPs. This removed some of these inefficiencies in communication that we anticipated would be a stimulus for Loop use. Furthermore, some iHCPs reported that they would want an intermediary to function in a similar administrative or facilitative role within Loop.

Other workflows relied on the patient (or family) to initiate communication, as with the transfer of information between organizations. In this situation, Loop's advantage in reducing the patient or caregiver's responsibility for transmitting information from HCP to HCP was not realized.

> *The patient is very helpful in communicating for herself. For example, she acts as the focal point for communication, prints out test results, and updates for me and provides them at the beginning of a visit for me to review.* [Fieldnote, iHCP, Site 1, month 3]

Overall Satisfaction With Using Loop

Monthly check-in interviews provided feedback on participants' satisfaction with Loop. Across all timepoints, 45 responses indicated users were "somewhat satisfied" to "very satisfied," and 7 responses indicated users "somewhat dissatisfied" and "very dissatisfied." Satisfaction feedback was only elicited if the participant had used the system in the previous month. We inferred that any interview that did not have a response for "satisfaction" or was coded as "unable to rate" had no Loop use. The denominator for the satisfaction question was 279 responses, and do not reflect independent responses because the same participants may have replied to this question at more than 1 timepoint.

COVID-19 Pandemic

The pandemic restrictions began in March 2020 in Canada and impacted recruitment and follow-up at Sites 4, 5, and 6. During this time, a parent caregiver would have liked to use Loop to check information about upcoming appointments, but at month 1, none of the additional HCPs had yet been assembled on their Loop. One iHCP reported loving the idea of Loop but felt it was difficult to build buy-in with other HCPs and patients during the pandemic. All contact with patients and other HCPs had shifted to virtual means and they found it hard to remember to talk about Loop. Another shift was that the pandemic led to removal of prior provincial restrictions on the use of alternative forms of communication and most care encounters became virtual. New billing codes for encounters by phone or videoconferencing were introduced. This change resulted in alternate methods of communication becoming incentivized and presented an unanticipated barrier to Loop use.

CFIR Interviews

CFIR interviews served to identify contextual barriers and facilitators according to this widely accepted determinant framework. CFIR comments were captured for each construct regarding its presence or absence in relation to supporting the implementation of Loop. Sites 2 and 6 had only 1 CFIR interview; sites 4 and 5 had 2 CFIR interviews; and sites 1 and 3 had 3 CFIR interviews. Notes for each site were summarized by construct by ES and RD and discussed with MB. Valence was rated for each construct by interview, but the mode could not be calculated for 2 sites where only 1 HCP was interviewed, and so valence is not reported numerically. Rather, coders reviewed the construct summaries by site and coded them as to whether they were perceived as supportive of implementation or not (yes/no), and as being present or absent in Loop implementation (+/-). Table 6 presents these findings: where codes were mixed, this is noted, and where constructs did not manifest in the interview, they are left blank.

Table 6. Salient CFIR[a] constructs by site.

CFIR domains and constructs	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
Interviews, n	1	3	3	1	2	2
Message frequency quartiles (Hi, Med, Lo)[b]	Q3 (Hi)	Q3 (Hi)	Q3 (Hi)	Q1 (Lo)	Q2 (Med)	Q2 (Med)
Intervention characteristics						
Intervention source	Yes (+)	Yes (+)	—	Yes (+)	Yes (+)	Yes (+)
Evidence strength and quality	—	—	Mixed	Yes (+)	Yes (+)	Yes (+)
Relative advantage	Yes (–)	Yes (+)	Mixed	Yes (–)	Yes (+)	Yes (+)
Adaptability	Yes (–)	Yes (–)	Yes (–)	Yes (–)	Yes (–)	Yes (–)
Complexity	Yes (–)	Yes (+)	Yes (+)	Yes (+)	Yes (+)	Yes (+)
Outer setting						
Patient needs and resources	Yes (+)	Yes (+)	Mixed	Yes (+)	Yes (+)	Yes (+)
Peer pressure	—	—	—	—	—	—
Cosmopolitanism (no score)	—	—	—	—	Yes (+)	—
External policies and incentives	Yes (+)	—	Yes (+)	—	Yes (+)	Yes (+)
Inner setting						
Structural characteristics	Yes (–)	Yes (–)	Yes (–)	Yes (–)	Yes (–)	Mixed
Networks and communications	—	—	Yes (–)	—	Yes (+)	Mixed
Culture	Yes (+)	Yes (+)	Mixed	Yes (–)	Yes (+)	Yes (+)
Implementation climate						
Tension for change	Yes (+)	Yes (+)	Mixed	—	Yes (–)	Yes (+)
Compatibility	Yes (–)	Yes (–)	Yes (–)	Yes (–)	Yes (–)	Yes (–)
Relative priority	—	Yes (–)	Mixed	—	Yes (–)	
Organizational incentives and rewards	—	—	Yes (–)	—	—	Yes (–)
Goals and feedback	—	—	Yes (–)	—	Yes (+)	—
Learning climate	Yes (–)	Yes (+)	Mixed	Yes (+)	Mixed	Yes (+)
Leadership engagement	Yes (–)	Yes (–)	Yes (–)	—	Yes (–)	
Available resources	Yes (+)	Yes (+)	Yes (+)	Yes (+)	Yes (+)	Yes (+)
Access to knowledge and information	Yes (+)	Yes (+)	Yes (+)	Yes (+)	Yes (+)	Yes (+)
Characteristics of individuals						
Knowledge and beliefs about the intervention	Yes (+)	Yes (+)	Mixed	Yes (+)	Yes (+)	Yes (+)
Self-efficacy	Yes (+)	Yes (+)	Yes (+)	Yes (+)	Yes (+)	Yes (+)
Individual stage of change (no score)	—	—	—	Yes (+)	Yes (+)	—
Individual identification with organization (no score)	—	Yes (+)	Mixed	Yes (+)	Yes (+)	—
Other personal attributes	Yes (+)	Yes (+)	Yes (+)	Yes (+)	Yes (+)	Yes (+)
Process						
Planning	Yes (+)	Yes (+)	Yes (+)	Yes (+)	Yes (+)	Mixed
Opinion leaders	—	Yes (+)	Mixed	Yes (+)	—	Mixed
Formally appointed internal implementation leaders	—	—	Yes (–)	Yes (+)	Mixed	—

CFIR domains and constructs	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
Champions	—	Yes (+)	Yes (–)	—	Yes (+)	—
External change agents	—	—	Yes (–)	—	Mixed	Yes (–)
Executing	—	—	Mixed	Yes (–)	Yes (+)	Mixed
Reflecting and evaluating	Yes (+)	Yes (+)	Yes (–)	—	Yes (+)	Yes (+)
Characteristics of recipients[c]						
Patient beliefs	Yes (+)	—	Mixed	Yes (+)	Yes (+)	Mixed
Patient experience	Yes (+)	Yes (+)	—	—	Yes (+)	Yes (–)

[a]CFIR: Consolidated Framework for Implementation Research.

[b]See Table 4.

[c]Not original to CFIR.

Several CFIR constructs were perceived as supporting implementation and as having manifested in Loop implementation across most sites. These are annotated as "Yes (+)" in Table 6.

Intervention Characteristics (Intervention Source and Complexity)

Respondents were aware of and had positive regard for where Loop originated, and this was perceived to be supportive of its implementation. Loop was viewed as easy to use, which was also facilitative for its implementation.

Outer Setting Characteristics (External Policies and Incentives)

Respondents perceived the outer health system context as supportive of tools that could improve communication within the patient's circle of care and saw this as supportive of Loop implementation.

Inner Setting Characteristics (Culture, Available Resources, Access to Knowledge, and Information)

Organizational culture was perceived as supportive of initiatives to implement evidence-based interventions such as Loop. Respondents felt they were well supported by the research team and had access to requisite knowledge and information about Loop in a way that supported implementation.

Characteristics of Individuals (Knowledge and Beliefs About the Intervention, Individual Identification With the Organization, Other Personal Attributes)

Respondents felt they were familiar with facts, truths, and principles related to Loop, perceived their organization was committed to evidence-based care, and possessed the requisite tolerance of ambiguity, intellectual ability, motivation, values, competence, capacity, and learning style.

Process (Planning, Reflecting, and Evaluating)

Respondents were aware of the plan in place to support Loop implementation and perceived this as facilitative. They also valued the opportunity to reflect on their experience and use of Loop during the monthly check-in interviews, although they were unsure as to how their reflections were used to inform ongoing implementation.

Characteristics of Recipients (Patient Beliefs and Patient Experience)

Respondents felt patients believed in the usefulness of Loop and that, for the most part, their experience of using Loop was positive.

Several CFIR constructs were perceived having the potential to support implementation generally but were perceived as absent with respect to Loop implementation across most sites—indicated as Yes (–) in Table 6.

Intervention Characteristics (Relative Advantage and Adaptability)

Respondents felt that although Loop had potential over alternative solutions, this relative advantage was not realized. They would have liked Loop to be adaptable to their workflow and environment, specifically with respect to its integration with the local EMR or patient portals.

Inner Setting Characteristics (Structural Characteristics, Tension for Change, Compatibility, Relative Priority, Organizational Incentives and Rewards, and Leadership Engagement)

With respect to the structural characteristics of the implementing organizations, respondents felt that Loop should ideally be integrated with the EMR they were already using. The lack of integration was perceived as a major barrier to Loop use as it meant users had to take several extra steps to enter and extract Loop information to put into the medical record.

> *Having more than one system to work with is always going to be very awkward.* [CFIR interview, iHCP, Site 1]

The lack of integration between Loop and EMRs also emerged as a barrier with respect to Loop's compatibility with existing workflows. These workflows are dictated by and facilitated through the EMR; anything outside of the EMR is difficult to manage.

> *It's already complicated by the fact that nurses who take care of the same patients, do not use the same record-keeping system as I do... so there is already additional work needed to fill this gap.* [CFIR interview, iHCP, Site 1]

Tension for change was low across all sites, given the perception that Loop was needed by only a small number of more complex patients. Respondents noted a lack of organizational incentives and rewards and limited involvement from organizational leadership. Several changes would be needed to implement Loop more effectively, including leadership engagement, use of reminders, and elimination of multiple passwords across various systems through integration with EMRs and portals.

Process (Formally Appointed Internal Implementation Leaders and External Change Agents)

Most of the constructs related to engaging others in the implementation process received mixed and largely negative ratings. Respondents commented on the limited influence of opinion leaders, champions, and external change agents.

We explored CFIR constructs found to be salient in this study relative to other studies in which CFIR was assessed by interview (see Table 7). Although 2 of the comparison studies [22,49] compared high and low implementing sites, 2 studies did not ([23] and this study). All 4 studies explored implementation in different contexts and for different interventions yet results for salient constructs are surprisingly similar in at least two or more of the studies, which suggest that several consistently robust constructs are commonly associated with implementation. Highly salient constructs across studies include relative advantage; patient needs and resources; external policies and incentives; tension for change; available resources; knowledge and beliefs about the intervention; and implementation planning. Salient to at least two studies were constructs of adaptability, complexity, structural characteristics, culture, compatibility, leadership engagement, access to information and knowledge, reflecting and evaluating, and beliefs of the health care recipient.

Table 7. Salient CFIR[a] domains across studies.

CFIR domains and constructs	Damschroder and Lowery (2013) [22][b]	Varsi et al (2015) [50][b]	Barwick et al (2015) [23][c]	This study (2021)[c]
Intervention characteristics				
Intervention source	—	—	—	Yes (+)
Evidence strength and quality	—	—	—	Yes (+)
Relative advantage[d]	Yes	Yes	Yes	Yes (–)
Trialability	—	Yes	—	—
Adaptability[e]	—	—	Yes	Yes (–)
Complexity[e]	—	—	Yes	Yes (+)
Outer setting				
Patient needs and resources[d]	Yes	Yes	Yes	Yes (+)
Cosmopolitanism	—	—	Yes	—
External policies and incentives[d]	Yes	—	Yes	Yes (+)
Inner setting				
Structural characteristics[e]	—	Yes	—	Yes (–)
Networks & Communications	Yes	—	—	—
Culture[e]	—	Yes	—	Yes (+)
Implementation climate				
Tension for change[d]	Yes	Yes	Yes	Yes (+)
Compatibility[e]	—	Yes	—	Yes (–)
Relative priority	Yes	Yes	—	—
Goals and feedback	Yes	—	—	—
Learning climate	Yes	—	—	—
Readiness for implementation				
Leadership engagement[e]	Yes	—	—	Yes (–)
Available resources[d]	Yes	Yes	—	Yes (+)
Access to information and knowledge[e]	—	—	Yes	Yes (+)
Characteristics of individuals				
Knowledge and beliefs about the intervention[d]	—	Yes	Yes	Yes (+)
Self-efficacy	—	—	—	Yes (+)
Individual identification with organization	—	—	—	Yes (+)
Other personal attributes	—	—	—	Yes (+)
Process				
Planning [d]	Yes	Yes	—	Yes (+)
Planning for sustainability	—	—	Yes	—
Opinion leaders	—	—	—	Yes (+)
Formally appointed internal implementation leaders	—	Yes	—	—
Champions	—	—	Yes	—

CFIR domains and constructs	Damschroder and Lowery (2013) [22][b]	Varsi et al (2015) [50][b]	Barwick et al (2015) [23][c]	This study (2021)[c]
Reflecting and evaluating[e]	Yes	—	—	Yes (+)
Characteristics of intervention recipients				
Patient beliefs[e]	—	—	Yes	Yes (+)
Patient experience	—	—	Yes	—

[a]CFIR: Consolidated Framework for Implementation Research.

[b]Related to constructs distinguishing between high and low implementers.

[c]Related to constructs identified as salient for implementation success: (+) construct was present; (–) construct was absent.

[d]Construct highly salient in more than 2 studies.

[e]Construct highly salient in at least two studies.

Clinical Outcomes

The primary outcome of COC was the Picker COC for which higher scores denote better COC. At baseline, the COC mean (SD) was 48.62 (29.88) for 48 patients. At month 2, the mean (SD) was 53.95 (29.18) for 19 patients. At month 6, the mean (SD) was 60.71 (26.79) for 14 patients. The descriptive statistics for all the surveys at each timepoint are presented in Table 8. COC scores showed a significant (P<.001) mean change (SD) of 24.23 (SD 26.01) in the positive direction at month 6 from baseline; however, none of the other timepoints showed a significant change in COC score and the rates of incomplete data limit any inference (Table 9). Similarly, an exploratory repeated measures analysis using generalized estimation method with autoregressive (AR-1) covariance structure for adjusting for repeated measures within patients with the outcome of COC, controlling for CCRQ and number of messages per Patient Loop, yielded no significant associations. We were unable to draw conclusions from the survey data with regard to appropriateness.

Table 8. Patient Surveys: Summary Descriptive Statistics.

Variable	Baseline, n/mean (SD)	M2, n/mean (SD)	M4, n/mean (SD)	M6, n/mean (SD)	M8, n/mean (SD)
COC[a]	48/48.62 (29.88)	19/53.95(29.18)	13/52.88 (24.02)	14/60.71 (26.79)	5/55.00 (22.71)
SSpedi[b]	6/22.50(12)	4/23.50 (4.12)	—	—	—
ESAS[c]					
Physical	40/17.02 (11.76)	16/16.25 (10.61)	13/16.54 (11.54)	14/17.71 (13.41)	5/17.00 (14.58)
Emotional	40/6.25 (5.93)	16/6.50 (6.34)	13/6.31 (6.34)	14/6.57 (6.93)	5/8.60 (7.02)
Well-being	40/4.33 (2.57)	16/4.44 (2.61)	13/4.62 (2.72)	14/4.21 (3.07)	5/4.20 (1.30)
Total symptom score	40/27.60 (17.65)	16/27.19 (17.21)	13/27.46 (18.80)	14/28.50 (22.60)	5/29.80 (20.29)
CCRQ[d]: CPDG[e]	44/81.25 (18.05)	19/83.77 (15.94)	12/77.15 (8.85)	13/82.40 (27.09)	5/95.00 (5.43)

[a]COC: Picker Ambulatory Cancer Care Scale, Continuity, and Coordination subscale.

[b]SSpedi: Symptom Screening in Pediatrics Tool.

[c]ESAS: Edmonton Symptom Assessment Scale.

[d]CCRQ: Client-Centered Rehabilitation Questionnaire.

[e]CPDG: Client Participation in Decision-making and Goal setting domain of CCRQ.

Table 9. Comparing patient baseline and follow-up COC[a] measurements.

Timepoint	N	Mean change in score (SD)	Comparison timepoints	P-value
M2	18	10.97 (36.76)	M2 to baseline	.246
M4	13	10.00 (41.50)	M4 to baseline	.402
M6	13	24.23 (26.01)	M6 to baseline	.006
M8	5	18.50 (20.36)	M8 to baseline	.112

[a]COC: Picker Ambulatory Cancer Care Scale, Continuity, and Coordination subscale.

Discussion

Principal Findings

This hybrid type II mixed methods implementation study found that gaps in communication and collaboration persist. In the absence of a shared clinical communication tool, health care providers have increasingly adopted email and texting with patients and caregivers over the last decade, and this virtualization has been accelerated by the COVID-19 pandemic [50-52]. Although participants acknowledged these forms of communication have their disadvantages and they could see potential in using Loop, this relative advantage was not realized largely due to the lack of Loop integration with existing health information systems (compatibility). There was a range of message frequency observed, with sites #1, #2 and #3 showing relatively more use than the others. These differences were not reflective of different construct profiles at the sites. Participants who used Loop were mostly satisfied. As in our previous study, patients were more likely than HCPs to initiate communication on Loop [8]. The implementation of Loop was done in the context of research, rather than as an organizational change initiative, and was therefore constrained in scope and time. The lack of broad organizational engagement, relative priority, and tension for change in the implementing organizations further hampered Loop's implementation. We were unable to make inferences about Loop's clinical effectiveness due to insufficient follow-up and survey data.

Implementation Effectiveness

In the context of this research implementation of Loop, our approach fell short in several ways and effective implementation remained elusive. Several studies have demonstrated that engaging local champions who sustain commitment and garner organizational support facilitates successful practice change [53-56]. Although we identified committed site champions, their role was informal. Engagement efforts largely fell to research staff and we did not seek to establish broad organizational support for this time-delimited implementation endeavor. This level of engagement was insufficient to influence greater adoption of Loop; however, sample size limits the conclusions we can draw. Moreover, it is likely that the lack of Loop compatibility and nonintegration with local EMRs would have proven to be a nonmodifiable barrier to implementation despite more active engagement.

Phase 1 activities and CFIR interviews affirmed that Loop was deemed acceptable at all sites. As in our previous pilot RCT [8], we demonstrated the feasibility of implementing Loop operationally, while once again experiencing challenges in enrolling additional HCPs, and finding the optimal patient and health care context (opportunity) to show the value of team-based communication.

Lack of compatibility and relative advantage were key implementation barriers in this study. Successful implementation of a novel tool, particularly one that disrupts [57] existing workflows, requires finding the optimal context and patient population to achieve early gains. HCPs frequently expressed that Loop ought to be integrated with the EMR used for charting. Indirect integration or workarounds that involved exporting messages as PDF and uploading them into the EMR were too cumbersome. Adding to this technological conundrum is that, in Ontario, as in many global jurisdictions, care teams straddle multiple EMRs even within the same organization. Many organizations must revert to custom nonscalable models for third-party tool integration. Although the landscape is shifting—for example, the 21st Century Cures legislation in the United States promotes greater standards adoption including HL7 FHIR [58]—this is far from the status quo in Canada. Furthermore, communication standards, as one might use for Loop, are not included in these health technologies, which are focused on the exchange of discrete data such as laboratory results, medications, or documents.

Some organizations have launched patient portals, but these too differ from one organization to another, perpetuating information silos. Of note, 2-way communication is not a standard feature of all patient portals. Patients and caregivers who had access to a patient portal felt that Loop would be more useful if it were integrated with their patient portal. This integration makes sense given that in both our Loop feasibility RCT and in this study, patients and caregivers were more often the drivers of communication. However, the prevailing institutionally tethered models may limit the addition of crucial external team members. Moreover, while patient portals with messaging capability have been shown to improve patient satisfaction and increase the "meaningful use" of data, few studies show that they improve health outcomes [59]. The nature of portals is changing with the emergence of institutionally agnostic commercial models such as Apple's Health Records entering the space, which may offer an opportunity for communication in the patient's circle of care.

Although participants perceived Loop's relative advantage over existing communication channels, this advantage was not actualized, presenting a key barrier [60-62] that would have been difficult to address using deimplementation strategies [63]. Patients, caregivers, and HCPs are reluctant to use yet another tool for communication. They would rather leverage Loop-like functionality (security, organized storage, and retrieval of clinical communication) in tools they already use to communicate in the nonhealth care aspects of their lives. However, none of the email platforms commonly in use provide these functions. Moving to a new communication tool will require realizing greater relative advantage, because efforts to deimplement commonly used means of communication are unlikely to work.

The Collaboration Space Model proposed by Eikey et al [5] outlines the following processes related to collaboration: workflow, communication, and information exchange [5]. The model also proposes 2 outcomes related to collaboration: maintaining awareness and establishing common ground. Applying this model to our study, we observed that participants who used Loop expressed that Loop had a mostly positive impact on the processes of communication and information exchange, and the outcome of maintaining awareness. Of note, a number of participants did not use Loop. The predominant perspectives were that Loop was disruptive of existing workflow. Although there were instances of coordination and cooperation occurring in Patient Loops, we did not observe an

impact on the higher-order collaboration outcome of establishing common ground. In addition to the barriers already discussed, our study was not designed to focus on the requirements for collaborative care in the context of a single site. In the absence of being able to effect change in organizational processes and structures, the adoption of Loop was hampered.

Clinical Effectiveness

Pooled survey data across all sites suggested an isolated improvement in continuity and coordination of care from baseline to month 6 but not at other timepoints. This should be interpreted with caution because fewer than 50% of participants completed survey responses beyond baseline.

Health System and Policy Environment

Since the start of the Loop research program in 2012, we have navigated an ever-shifting landscape. Ontario's health care environment has experienced major policy changes in the past 3 years, including the restructuring of regional health authorities (Local Health Integration Networks) into a new model of networks of care called Ontario Health Teams. While this new organizational structure may hold promise for the scaling of eHealth solutions in the future, the transitional period has resulted in deferred decision making. Although we cannot be certain, these system changes may have impacted the progress of our work in terms of informing policy and partnering with provincial organizations for the sustainability of Loop beyond the research program.

A relevant policy shift emerging from the COVID-19 pandemic was the necessity of providing care outside the in-person visit, and permission to communicate and bill for care using phone and video. The perspectives that emerged during our study suggest that the net effect of this system shift was unfavorable for Loop use.

A larger structural issue is related to models of compensation for physicians, nurses, and other HCPs that incentivize in-person or synchronous virtual encounters. Physician fee for service, capitated, and alternate payment plans likely impact the use of Loop. Notably a recent study found very high rates of patient and provider desire to engage in asynchronous messaging preferentially versus using synchronous video or phone in a primary care [64], capitated setting where the family physician and patients accrue benefits without penalties. For physicians working in the fee for service model, including most specialists, there are no billing codes for asynchronous communication. Similarly, there are different models of compensation for nurses. As an example, in Ontario, home care nurses are required to do a certain number of face-to-face visits under one model. The nurses who participated in our study were salaried. All iHCPs were affiliated with academic organizations, likely influencing their willingness to participate in research. We were unable to observe a difference in Loop adoption across different forms of HCP compensation.

Some health system changes were facilitative for Loop. In 2019, the College of Physicians and Surgeons of Ontario disseminated 4 interrelated policies for improving COC in Ontario [65]. While this action raised COC as a priority, it focused mainly on ensuring that physicians provide appropriate options for after-hours care and reliable processes for effective transitions in care. It is possible that compliance with these recommendations will advantage use of asynchronous team-based communication. However, these recommendations have been met with resistance from various stakeholders and their implementation and long-term impact remain uncertain.

Future Research, Health Equity, and Accessibility in eHealth Tools

The challenges encountered in this study are common in studies of eHealth tools, and more generally of implementing complex interventions in complex settings. However, a compelling finding was that Loop disrupted workflows and workarounds that have developed in the absence of standardized tools for communication. Loop could not transcend established communication modalities despite their inability to enable team communication and collaboration. Future research could specifically focus on a health care region, integrate Loop with existing eHealth tools, pilot a compensation structure for asynchronous communication, deimplement or re-design optimal communication workflows from the ground up, and support requirements or conditions for collaboration to occur. In addition, future research should examine the impact of eHealth implementation on health disparities, which have been shown to increase with the introduction of patient–provider messaging tools [66]. Despite the considerable and seemingly intractable barriers identified in this study, the trend toward more digital communication in health care is inevitable and it is likely that an interoperable system of communication and documentation will emerge in time. It is important that tools such as Loop be available and accessible to all, so that health inequities are not further magnified.

Limitations

The research context of this implementation endeavor likely introduced bias insofar as health care providers and patients were possibly more inclined to use Loop to fulfill their commitment as research participants. In addition to participant, team, and site selection bias, we acknowledge the possible researcher bias: the main reviewer (AH) for the qualitative check-in analysis is also the lead for the Loop research program.

Recruitment may have been impacted by issues of equity and access. Although smartphone and internet penetration are rising in Ontario, increasing from 81.4% in 2010 [67] to 92.2% in 2018 [68] with 89.1% of Ontarians reporting having a smartphone for personal use in 2018 [69], there are persisting disparities in access among the rural and marginally housed [70,71]. Given that internet access is a core component of Loop, access is not equitable to all potential patients.

Beyond the anticipated patient attrition rates, there was an unexpectedly low completion rate for clinical survey data beyond baseline. This limited the inferences that could be made about Loop's clinical effectiveness. Participants from whom we were unable to collect data may have been more likely to talk about barriers to using Loop, and this too may have skewed our qualitative analyses. The challenges in recruiting participants at each of the sites limited our ability to mitigate attrition by over-recruiting within the timelines of the study. As in the

previous study, recruiting additional HCPs to participate on the patient's Loop team remained difficult and limited the use of Loop among those already enrolled.

The COVID-19 pandemic resulted in dramatic changes to the health care system in Ontario starting in March 2020, impacting recruitment, relative advantage in light of encouraged use of phone and email and permissible billing, and data collection in the follow-up periods (Phases 2 and 3) at Sites 4, 5, and 6. In the first 2 months of the pandemic, all research study recruitment in Ontario was paused unless deemed essential to the health of the participant or relevant to the pandemic. Although the impact was felt on many levels including study staff workflow, the main challenge was that the attention of many HCPs and health care leaders was focused on planning for the health care challenges posed by the pandemic.

The study was limited in being able to support the collaborative requirements at each site. Therefore, a weakness of the study was its focus on Loop without being able to substantively support the structures and processes that would have allowed us to impact collaboration.

Conclusions

This study highlighted the importance of system and organizational context and several key determinants of effective implementation. From the start of the Loop program of research, regulatory guidelines have restricted the use of email and text due to privacy concerns and created data silos within organizations. Despite these restrictions and in the absence of other practical tools for communication, there has been a steady increase in the use of email, text, and other forms of messaging to provide the care that patients need. The COVID-19 pandemic shed light on the essential components of that care. Delivering care became the priority and the regulatory guidelines became pragmatic. The health system learned that a considerable proportion of the care HCPs provide in person can be provided virtually, by video or phone, if the HCP is compensated for this mode of service delivery. If a rational approach to the regulatory framework continues and health leadership prioritizes an integrated digital infrastructure, we may yet achieve the goal of care being provided by the right person, in the right place, at the right time, and with the right tools.

Key facilitative factors again show themselves to be essential for effective implementation. Perceived relative advantage only goes so far, and this study demonstrates, yet again, that compatibility, relative advantage, tension for change, and engagement are essential implementation components that must be realized.

Fundamental structural challenges remain for the implementation and scaling of a shared system of asynchronous communication, including digital integration and a fee structure for compensation. If a new hybrid model of care emerges from the pandemic, it is likely to de-emphasize in-person encounters between patients and HCPs, disrupt existing workflows, and allow the intentional design of new pathways for care that prioritize team communication, access, and COC. Until these changes manifest, effective implementation of Loop and similar communication platforms will continue to be elusive.

Acknowledgments

This research was funded by the Canadian Cancer Society (Innovation to Impact grant #704797). The Temmy Latner Centre for Palliative Care, Sinai Health, provided in-kind support. We wish to acknowledge additional members of the Loop Research Collaborative: Joe Cafazzo, Peter Weinstein, Monika Krzyżanowska, Peter Selby, Andrea Bezjak, Mayura Loganathan, and Russell Goldman. Additionally, we are grateful for the work and support of the following individuals: Maya Watson, Amy Clarke, Rahim Moineddin, Jodheme Goldhar, and the implementation champions and health care providers at each of our recruitment sites.

Authors' Contributions

Except for the first author (AH [principal]) and the last author (MB [senior]), all coauthors are listed in alphabetical order.

Conflicts of Interest

AH has submitted an invention disclosure application for Loop to the Sinai Health Technology Transfer office; however, no commercialization activity has been undertaken. There is an intellectual property agreement with regard to Loop between the collaborating organizations, Sinai Health, University Health Network, and Sick Children's Hospital. Other authors have nothing to declare.

Multimedia Appendix 1
Standards for Reporting Implementation Studies: the StaRI checklist for completion.
[PDF File (Adobe PDF File), 249 KB - jmir_v23i3e25505_app1.pdf]

Multimedia Appendix 2
Semi-structured interview for Health-Care Providers based on the Consolidated Framework for Implementation Research.
[PDF File (Adobe PDF File), 277 KB - jmir_v23i3e25505_app2.pdf]

References

1. Bates DW. Health Information Technology and Care Coordination: The Next Big Opportunity for Informatics? Yearb Med Inform 2015 Aug 13;10(1):11-14 [FREE Full text] [doi: 10.15265/IY-2015-020] [Medline: 26123913]
2. Ministry of Health and Long-Term Care. Preventing and Managing Chronic Disease: Ontario's Framework. Ontario, Canada: Ministry of Health and Long-Term Care; 2007. URL: http://www.health.gov.on.ca/en/pro/programs/cdpm/pdf/framework_full.pdf [accessed 2021-02-08]
3. Craig C, Eby D, Whittington J. Coordination Model: Better Care at Lower Cost for People with Multiple Health and Social Needs (IHI Innovation Series White Paper). Cambridge, Massachusetts: Institute for Health Care Improvement; 2011. URL: http://www.ihi.org/resources/Pages/IHIWhitePapers/IHICareCoordinationModelWhitePaper.aspx [accessed 2021-02-18]
4. Fuks H, Raposo A, Gerosa MA, Pimental, Mariano. The 3C collaboration model. In: Encyclopedia of E-Collaboration. Hershey, PA: IGI Global; 2008:637-644.
5. Eikey EV, Reddy MC, Kuziemsky CE. Examining the role of collaboration in studies of health information technologies in biomedical informatics: A systematic review of 25 years of research. J Biomed Inform 2015 Oct;57:263-277 [FREE Full text] [doi: 10.1016/j.jbi.2015.08.006] [Medline: 26264406]
6. Haggerty JL, Reid RJ, Freeman GK, Starfield BH, Adair CE, McKendry R. Continuity of care: a multidisciplinary review. BMJ 2003 Nov 22;327(7425):1219-1221 [FREE Full text] [doi: 10.1136/bmj.327.7425.1219] [Medline: 14630762]
7. National Research Corporation. Development and Validation of the Picker Ambulatory Oncology Survey Instrument in Canada. Lincoln: Ontario Hospital Association; 2003.
8. Voruganti T, Grunfeld E, Jamieson T, Kurahashi AM, Lokuge B, Krzyzanowska MK, et al. My Team of Care Study: A Pilot Randomized Controlled Trial of a Web-Based Communication Tool for Collaborative Care in Patients With Advanced Cancer. J Med Internet Res 2017 Jul 18;19(7):e219 [FREE Full text] [doi: 10.2196/jmir.7421] [Medline: 28720558]
9. Jamieson T. Using social network analysis to analyze the function of collaborative care teams in an online social networking tool. Portland, OR: Oregon Health and Science University; 2003. URL: https://digitalcollections.ohsu.edu/concern/etds/4m90dv71h?locale=es [accessed 2021-02-18]
10. Jamieson T, Husain A, Daub A, Goldhar A. Us and Them: Why patient-defined health teams don't communicate well. Healthy Debates. 2016. URL: http://healthydebate.ca/opinions/us-and-them-why-patient-defined-health-teams-dont-communicate-well [accessed 2021-02-18]
11. Activate Care [software]. 2012. URL: https://www.activatecare.com/about-activate [accessed 2021-02-18]
12. Trillian [software]. 2000. URL: https://trillian.im/hipaa-compliant-messaging/ [accessed 2021-02-08]
13. Hucu.ai [software]. 2019. URL: https://www.hucu.ai/how-it-works/ [accessed 2021-02-08]
14. Halo Health [software]. 2010. URL: https://www.halohealth.ca/about-us [accessed 2021-02-08]
15. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
16. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
17. Glasgow RE, Phillips SM, Sanchez MA. Implementation science approaches for integrating eHealth research into practice and policy. Int J Med Inform 2014 Jul;83(7):e1-11 [FREE Full text] [doi: 10.1016/j.ijmedinf.2013.07.002] [Medline: 23910896]
18. Stange KC, Glasgow RE. Contextual Factors: The Importance of Considering and Reporting on Context in Research on the Patient-Centered Medical Home. Rockville, MD: Agency for Healthcare Research and Quality; 2013:13-0045-EF.
19. Glasgow R, Green LW, Taylor MV, Stange KC. An evidence integration triangle for aligning science with policy and practice. Am J Prev Med 2012 Jun;42(6):646-654 [FREE Full text] [doi: 10.1016/j.amepre.2012.02.016] [Medline: 22608384]
20. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
21. Meyers DC, Durlak JA, Wandersman A. The quality implementation framework: a synthesis of critical steps in the implementation process. Am J Community Psychol 2012 Dec;50(3-4):462-480. [doi: 10.1007/s10464-012-9522-x] [Medline: 22644083]
22. Damschroder LJ, Lowery JC. Evaluation of a large-scale weight management program using the consolidated framework for implementation research (CFIR). Implement Sci 2013 May 10;8:51 [FREE Full text] [doi: 10.1186/1748-5908-8-51] [Medline: 23663819]
23. Barwick M, Barac R, Zlotkin S. Implementation of Exclusive Breastfeeding in Ethiopia and Mali: Factors Influencing Change. CARE Canada, Save the Children Canada, Plan International Canada. Ontario, Canada: Muskoka Initiative Consortium URL: http://melaniebarwick.com/wp-content/uploads/dlm_uploads/2019/02/EBF-Research-Report-FINAL-July-29-2015.pdf [accessed 2021-02-18]

24. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
25. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, Griffiths CJ, StaRI Group. Standards for Reporting Implementation Studies (StaRI) Statement. BMJ 2017 Mar 06;356:i6795 [FREE Full text] [doi: 10.1136/bmj.i6795] [Medline: 28264797]
26. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
27. Kurahashi AM, Weinstein PB, Jamieson T, Stinson JN, Cafazzo JA, Lokuge B, et al. In the Loop: The Organization of Team-Based Communication in a Patient-Centered Clinical Collaboration System. JMIR Hum Factors 2016 Mar 24;3(1):e12 [FREE Full text] [doi: 10.2196/humanfactors.4996] [Medline: 27025912]
28. Kurahashi AM, Stinson JN, van Wyk M, Luca S, Jamieson T, Weinstein P, et al. The Perceived Ease of Use and Usefulness of Loop: Evaluation and Content Analysis of a Web-Based Clinical Collaboration System. JMIR Hum Factors 2018 Jan 09;5(1):e2 [FREE Full text] [doi: 10.2196/humanfactors.7882] [Medline: 29317386]
29. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982 Dec;5(6):649-655. [Medline: 7165009]
30. ECOG-ACRIN Cancer Research Group. ECOG Performance Status. 2020 Nov. URL: https://ecog-acrin.org/resources/ecog-performance-status [accessed 2021-02-18]
31. Stillman MJ, Rybicki LA. The bedside confusion scale: development of a portable bedside test for confusion and its application to the palliative medicine population. J Palliat Med 2000;3(4):449-456. [doi: 10.1089/jpm.2000.3.4.449] [Medline: 15859697]
32. Fixsen D, Naoom S, Blase K, Friedman R, Wallace F. Implementation Research: A Synthesis of the Literature (FMHI 231). Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network; 2005. URL: https://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/NIRN-MonographFull-01-2005.pdf [accessed 2021-02-18]
33. Implementation Research: A Synthesis of the Literature. 2018. URL: https://nirn.fpg.unc.edu/resources/hexagon-exploration-tool [accessed 2021-02-08]
34. Downing G. Q&A Manual, Instructions & Definitions for Use of Palliative Performance Scale (PPSv2). Victoria, Canada: Victoria Hospice Society; 2020. URL: https://victoriahospice.org/wp-content/uploads/2020/08/PPSv2-QA-Instructions-and-Definitions-updated-July-2020.pdf [accessed 2021-02-08]
35. Yang CC, Chen PC, Hsu CW, Chang SL, Lee CC. Validity of the age-adjusted charlson comorbidity index on clinical outcomes for patients with nasopharyngeal cancer post radiation treatment: a 5-year nationwide cohort study. PLoS One 2015;10(1):e0117323 [FREE Full text] [doi: 10.1371/journal.pone.0117323] [Medline: 25617629]
36. Schaink AK, Kuluski K, Lyons RF, Fortin M, Jadad AR, Upshur R, et al. A scoping review and thematic classification of patient complexity: offering a unifying framework. J Comorb 2012;2:1-9 [FREE Full text] [doi: 10.15256/joc.2012.2.15] [Medline: 29090137]
37. Torres-Espíndola LM, Demetrio-Ríos J, Carmona-Aparicio L, Galván-Díaz C, Pérez-García M, Chávez-Pacheco JL, et al. Comorbidity Index as a Predictor of Mortality in Pediatric Patients With Solid Tumors. Front Pediatr 2019 Jan 24;7(1):48 [FREE Full text] [doi: 10.3389/fped.2019.00048] [Medline: 30881949]
38. Mitchell P, Wynia M, Golden R, McNellis B, Okun S, Webb CE, et al. Core Principles & Values of Effective Team-based Health Care. Washington, DC: National Academy of Medicine; 2012.
39. Bruera E, Kuehn N, Miller MJ, Selmser P, Macmillan K. The Edmonton Symptom Assessment System (ESAS): a simple method for the assessment of palliative care patients. J Palliat Care 1991;7(2):6-9. [Medline: 1714502]
40. Hui D, Bruera E. The Edmonton Symptom Assessment System 25 Years Later: Past, Present, and Future Developments. J Pain Symptom Manage 2017 Mar;53(3):630-643 [FREE Full text] [doi: 10.1016/j.jpainsymman.2016.10.370] [Medline: 28042071]
41. Dupuis LL, Johnston DL, Baggott C, Hyslop S, Tomlinson D, Gibson P, et al. Validation of the Symptom Screening in Pediatrics Tool in Children Receiving Cancer Treatments. J Natl Cancer Inst 2018 Jun 01;110(6):661-668 [FREE Full text] [doi: 10.1093/jnci/djx250] [Medline: 29272441]
42. Johnston DL, Hyslop S, Tomlinson D, Baggott C, Gibson P, Orsey A, et al. Describing symptoms using the Symptom Screening in Pediatrics Tool in hospitalized children with cancer and hematopoietic stem cell transplant recipients. Cancer Med 2018 May;7(5):1750-1755 [FREE Full text] [doi: 10.1002/cam4.1433] [Medline: 29573197]
43. Cott CA, Teare G, McGilton KS, Lineker S. Reliability and construct validity of the client-centred rehabilitation questionnaire. Disabil Rehabil 2006 Nov 30;28(22):1387-1397. [doi: 10.1080/09638280600638398] [Medline: 17071570]
44. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
45. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]

46. Gale RC, Wu J, Erhardt T, Bounthavong M, Reardon CM, Damschroder LJ, et al. Comparison of rapid vs in-depth qualitative analytic methods from a process evaluation of academic detailing in the Veterans Health Administration. Implement Sci 2019 Feb 01;14(1):11-88 [FREE Full text] [doi: 10.1186/s13012-019-0853-y] [Medline: 30709368]
47. Barwick M, Dubrowski R, Damschroder L. Factors Associated with Effective Implementation: Research and Practical Implications. In: Albers B, Shlonsky A, Mildon R, editors. Implementation Science 3.0. Cham, Switzerland: Springer; 2020:398-405.
48. Creswell JW. Research Design: Quantitative and Mixed Methods Approaches (2nd ed). Thousand Oaks, CA: Sage Publications; 2003.
49. Varsi C, Ekstedt M, Gammon D, Ruland CM. Using the Consolidated Framework for Implementation Research to Identify Barriers and Facilitators for the Implementation of an Internet-Based Patient-Provider Communication Service in Five Settings: A Qualitative Study. J Med Internet Res 2015 Nov 18;17(11):e262 [FREE Full text] [doi: 10.2196/jmir.5091] [Medline: 26582138]
50. Ontario Health. Recommendations for Regional Health Care Delivery During the COVID-19 Pandemic: Outpatient Care, Primary Care, and Home and Community Care. 2020. URL: https://www.wrh.on.ca/uploads/Coronavirus/Recommendations_Regional_Health_Care_Delivery_COVID19_June_8_2020.pdf [accessed 2021-02-18]
51. Canadian Medical Association. What Canadians Think About Virtual Health Care. 2020. URL: https://www.cma.ca/news/virtual-care-real-care-national-poll-shows-canadians-are-overwhelmingly-satisfied-virtual [accessed 2021-02-18]
52. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
53. Bunce AE, Gruß I, Davis JV, Cowburn S, Cohen D, Oakley J, et al. Lessons learned about the effective operationalization of champions as an implementation strategy: results from a qualitative process evaluation of a pragmatic trial. Implement Sci 2020 Oct 01;15(1):87 [FREE Full text] [doi: 10.1186/s13012-020-01048-1] [Medline: 32998750]
54. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004 Apr 20;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
55. Helfrich CD, Weiner BJ, McKinney MM, Minasian L. Determinants of implementation effectiveness: adapting a framework for complex innovations. Med Care Res Rev 2007 Jun;64(3):279-303. [doi: 10.1177/1077558707299887] [Medline: 17507459]
56. Turner R. Diffusion of innovations [Book review]. Journal of Minimally Invasive Gynecology 2007 Nov;14(6):776-629 [FREE Full text] [doi: 10.1016/j.jmig.2007.07.001]
57. Voruganti T, Husain A, Grunfeld E, Webster F. Disruption or innovation? A qualitative descriptive study on the use of electronic patient-physician communication in patients with advanced cancer. Support Care Cancer 2018 Aug;26(8):2785-2792. [doi: 10.1007/s00520-018-4103-7] [Medline: 29502156]
58. 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program. Washington, DC: Office of the National Coordinator for Health Information Technology, Department of Health and Human Services; 2020. URL: https://www.healthit.gov/sites/default/files/cures/2020-03/ONC_Cures_Act_Final_Rule_03092020.pdf [accessed 2021-02-18]
59. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
60. Tucker AL, Nembhard IM, Edmondson AC. Implementing New Practices: An Empirical Study of Organizational Learning in Hospital Intensive Care Units. Management Science 2007 Jun;53(6):894-907 [FREE Full text] [doi: 10.1287/mnsc.1060.0692]
61. Grol RPTM, Bosch MC, Hulscher MEJL, Eccles MP, Wensing M. Planning and studying improvement in patient care: the use of theoretical perspectives. Milbank Q 2007;85(1):93-138 [FREE Full text] [doi: 10.1111/j.1468-0009.2007.00478.x] [Medline: 17319808]
62. Feldstein AC, Glasgow RE. A Practical, Robust Implementation and Sustainability Model (PRISM) for Integrating Research Findings into Practice. The Joint Commission Journal on Quality and Patient Safety 2008 Apr;34(4):228-243. [doi: 10.1016/s1553-7250(08)34030-6]
63. Wang V, Maciejewski ML, Helfrich CD, Weiner BJ. Working smarter not harder: Coupling implementation to de-implementation. Healthc (Amst) 2018 Jun;6(2):104-107 [FREE Full text] [doi: 10.1016/j.hjdsi.2017.12.004] [Medline: 29279297]
64. Women's College Hospital Institute for Health Systems Solutions and Virtual Care. Enhanced Access to Primary Care: Project Evaluation Final Report. Ontario Telemedicine Network The Ministry of Health and Long-Term Care: Toronto, ON; 2019. URL: https://otn.ca/wp-content/uploads/2019/08/eapc-evaluation-report.pdf [accessed 2021-02-18]
65. Continuity of Care Policies. Toronto, ON: College of Physicians and Surgeons of Ontario; 2019. URL: https://www.cpso.on.ca/Physicians/Policies-Guidance/Policies/Continuity [accessed 2021-02-18]
66. Senft N, Butler E, Everson J. Growing Disparities in Patient-Provider Messaging: Trend Analysis Before and After Supportive Policy. J Med Internet Res 2019 Oct 07;21(10):e14976 [FREE Full text] [doi: 10.2196/14976] [Medline: 31593539]

67. Household Access to the Internet at Home, by Household Income Quartile and Geography. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2210000701 [accessed 2021-02-16]
68. Internet Use by Province. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2210008301 [accessed 2021-02-18]
69. Use of Internet Services and Technologies by Age Group and Household Income Quartile. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2210011301 [accessed 2020-09-16]
70. Haight M, Quan-Haase A, Corbett BA. Revisiting the digital divide in Canada: the impact of demographic factors on access to the internet, level of online activity, and social networking site usage. Information, Communication & Society 2014 Mar 07;17(4):503-519 [FREE Full text] [doi: 10.1080/1369118X.2014.891633]
71. Communications Monitoring Report. Ottawa, Canada: Canadian Radio-television and Telecommunications Commission; 2019. URL: https://crtc.gc.ca/pubs/cmr2019-en.pdf [accessed 2021-02-16]

Abbreviations

ACCI: Age-Adjusted Charlson Comorbidity Index
CCRQ: Client-Centered Rehabilitation Questionnaire
CPDG: Client Participation in Decision Making domain of CCRQ
CFIR: Consolidated Framework for Implementation Research
COC: continuity of care
ECOG: Eastern Cooperative Oncology Group
EMR: electronic medical record
HCP: health care provider
iHCP: initiating health care provider
QIF: Quality Improvement Framework
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 25.11.20; peer-reviewed by C Kuziemsky; comments to author 16.12.20; revised version received 24.12.20; accepted 22.01.21; published 03.03.21.

Please cite as:
Husain A, Cohen E, Dubrowski R, Jamieson T, Kurahashi AM, Lokuge B, Rapoport A, Saunders S, Stasiulis E, Stinson J, Subramaniam S, Wegier P, Barwick M
A Clinical Communication Tool (Loop) for Team-Based Care in Pediatric and Adult Care Settings: Hybrid Mixed Methods Implementation Study
J Med Internet Res 2021;23(3):e25505
URL: https://www.jmir.org/2021/3/e25505
doi:10.2196/25505
PMID:33656445

Original Paper

An 11-Item Measure of User- and Human-Centered Design for Personal Health Tools (UCD-11): Development and Validation

Holly O Witteman[1,2,3], PhD; Gratianne Vaisson[1,3], MSc; Thierry Provencher[1], BA; Selma Chipenda Dansokho[1], PhD; Heather Colquhoun[4], PhD, OT Reg (Ont); Michele Dugas[1,2], MSc; Angela Fagerlin[5,6], PhD; Anik MC Giguere[1,2], PhD; Lynne Haslett[7], BSCN, MEd, NP-PHC; Aubri Hoffman[8], PhD; Noah M Ivers[4,9], MD, PhD; France Légaré[1,2], MD, PhD; Marie-Eve Trottier[1,3], MSc; Dawn Stacey[10,11], RN, PhD; Robert J Volk[8], PhD; Jean-Sébastien Renaud[1,2], PhD

[1]Université Laval, Quebec City, QC, Canada
[2]VITAM Research Centre for Sustainable Health, Quebec City, QC, Canada
[3]CHU de Québec-Université Laval, Quebec City, QC, Canada
[4]University of Toronto, Toronto, ON, Canada
[5]University of Utah, Salt Lake City, UT, United States
[6]Salt Lake City VA Center for Informatics Decision Enhancement and Surveillance, Salt Lake City, UT, United States
[7]East End Community Health Centre, Toronto, ON, Canada
[8]The University of Texas MD Anderson Cancer Center, Houston, TX, United States
[9]Women's College Hospital, Toronto, ON, Canada
[10]University of Ottawa, Ottawa, ON, Canada
[11]Ottawa Hospital Research Institute, Ottawa, ON, Canada

Corresponding Author:
Holly O Witteman, PhD
Université Laval
1050, avenue de la Médecine
Quebec City, QC
Canada
Phone: 1 418 656 2131
Email: holly.witteman@fmed.ulaval.ca

Abstract

Background: Researchers developing personal health tools employ a range of approaches to involve prospective users in design and development.

Objective: The aim of this paper was to develop a validated measure of the human- or user-centeredness of design and development processes for personal health tools.

Methods: We conducted a psychometric analysis of data from a previous systematic review of the design and development processes of 348 personal health tools. Using a conceptual framework of user-centered design, our team of patients, caregivers, health professionals, tool developers, and researchers analyzed how specific practices in tool design and development might be combined and used as a measure. We prioritized variables according to their importance within the conceptual framework and validated the resultant measure using principal component analysis with Varimax rotation, classical item analysis, and confirmatory factor analysis.

Results: We retained 11 items in a 3-factor structure explaining 68% of the variance in the data. The Cronbach alpha was .72. Confirmatory factor analysis supported our hypothesis of a latent construct of user-centeredness. Items were whether or not: (1) patient, family, caregiver, or surrogate users were involved in the steps that help tool developers understand users or (2) develop a prototype, (3) asked their opinions, (4) observed using the tool or (5) involved in steps intended to evaluate the tool, (6) the process had 3 or more iterative cycles, (7) changes between cycles were explicitly reported, (8) health professionals were asked their opinion and (9) consulted before the first prototype was developed or (10) between initial and final prototypes, and (11) a panel of other experts was involved.

Conclusions: The User-Centered Design 11-item measure (UCD-11) may be used to quantitatively document the user/human-centeredness of design and development processes of patient-centered tools. By building an evidence base about such processes, we can help ensure that tools are adapted to people who will use them, rather than requiring people to adapt to tools.

(J Med Internet Res 2021;23(3):e15032) doi:10.2196/15032

KEYWORDS

patient-centered care; patient participation; health services research; validation studies as topic; surveys and questionnaires; humans; user-centred design, human-centred design; user-centered design; human-centered design; co-design; instrument; scale; index; patient and public involvement

Introduction

Many products and applications aim to support people in managing their health and living their lives. These include physical tools like wheelchairs [1] or eating utensils [2], medical devices like insulin pumps [3] or home dialysis equipment [4], assistive devices like screen readers [5] or voice aids [6], digital applications like eHealth tools [7] or mHealth (mobile health) tools [8,9], tools for collecting patient-reported outcome or experience measures [10,11], patient decision aids [12], and a variety of other personal health tools.

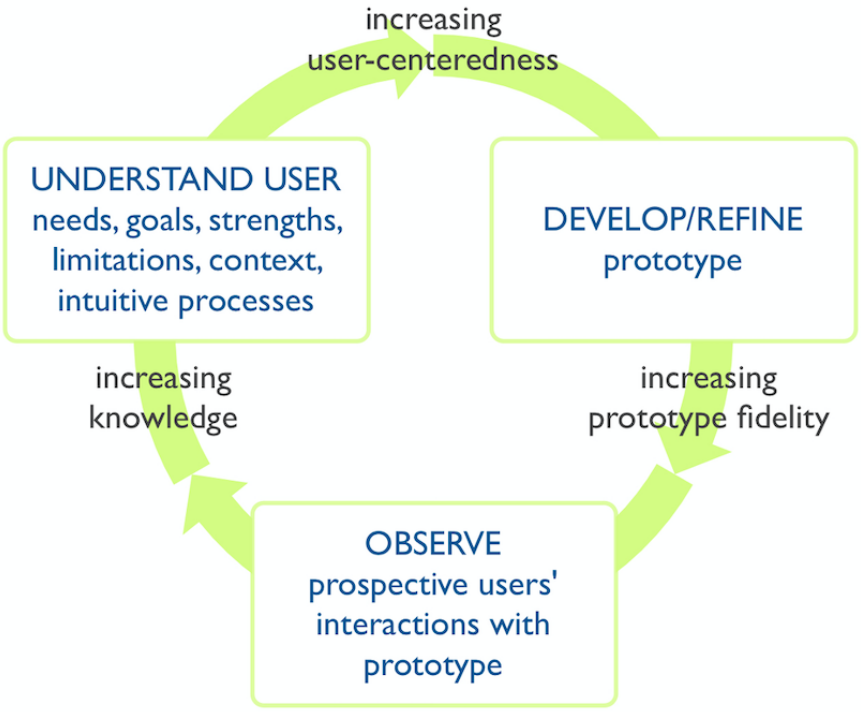
None of these tools can achieve their intended impact if they are not usable by and useful to their intended users. Accordingly, designers and developers frequently seek to involve users in design and development processes to ensure such usability and utility. In a previous systematic review of the design and development processes of a range of personal health tools, we documented that the extent and type of user involvement varies widely [13]. Structured ways to describe this variation could help capture data across projects and may serve to build an evidence base about the potential effects of design and development processes.

The systematic review was grounded in a framework of user-centered design [14], shown in Figure 1, that we had synthesized from foundational literature. In this framework, a user is any person who interacts with (in other words, uses) a system, service, or product for some purpose. User-centered design is a long-standing approach [15], sometimes referred to as human-centered design [16], that is both conceptually and methodologically related to terms like design thinking and co-design [17]. It is intended to optimize the user experience of a system, service, or product [18-21]. While user-centered design is not the only approach that may facilitate such optimization, it served as a useful overall framework for structuring the data reported in the papers included in our systematic review. In our work, we define user-centered design as a fully or semistructured approach in which people who currently use or who could in the future use a system, service, or product are involved in an iterative process of optimizing its user experience. This iterative process includes one or more steps to understand prospective users, including their needs, goals, strengths, limitations, contexts (eg, the situations or environments in which they will use a tool), and intuitive processes (eg, the ways in which they currently address the issue at hand or use similar systems, services, or products). The iterative process also includes one or more steps to develop or refine prototypes, and one or more steps to observe prospective users' interactions with versions of the tool.

Iivari and Iivari [22] noted that the different ways in which user-centeredness is described in the literature imply four distinct meanings or dimensions: (1) user focus, meaning that the system is designed and developed around users' needs and capabilities; (2) work-centeredness, meaning that the system is designed and developed around users' workflow and tasks; (3) user involvement or participation, meaning that the design and development process involves users or users participate in the process; and (4) system personalization, meaning the system is individualized by or for individual users. Our definition of user-centeredness and framework of user-centered design draw most strongly upon the third of these (user involvement or participation) as a means to achieve the first (user focus) and fourth (system personalization). The second meaning (work-centeredness) is less relevant here as it refers to paid work in the original definition. However, it may be worth noting the considerable work that people may need to undertake to make health decisions or to live with illness or disability [23-26].

In our previous systematic review, we used the above framework of user-centered design to extract and organize the extracted data from 623 articles describing the design, development, or evaluation processes of 390 personal health tools, predominantly patient decision aids, which are tools intended to support personal health decisions [13]. We documented a wide range of practices, leading us to question whether it might be possible to use this structured data set to develop a measure to capture aspects of the user-centeredness of design and development processes, similar to how other measures capture complex concepts or processes that are not directly observable; for example, social capital [27], learning processes [28], health-related quality of life [29], and health care quality [30]. We posited that although a high-level summary of design and development processes would not be able to capture nuances within each project, it may nonetheless be valuable to be able to capture information that would otherwise be difficult to synthesize across diverse projects. Of the 390 included personal health tools in our previous systematic review, 348 met our prespecified criterion regarding sufficient information related to the design and development processes, while the other 42 reported information only about their evaluation. Therefore, in this study, using an existing structured data set describing the design and development of 348 personal health tools, we aimed to derive a measure of the user- or human-centeredness of the design and development of personal health tools.

Figure 1. User-centered design framework.



Methods

Validity Framework and Overall Approach

Guided by an established validity framework [31], we developed and validated a measure using classical test theory. Classical test theory is a set of concepts and methods developed over decades [32-35] based on the earlier work of Spearman [36,37]. It posits that it is possible to develop items that each assess part of a construct that we wish to measure but is not directly observable; for example, patient-reported outcomes [38,39], responsibility and cooperation in a group learning task [40], or, in our case, the user-centeredness of a design and development process. Classical test theory further posits that each item captures part of what we wish to measure, plus error, and assumes that the error is random. This means that as the number of items increases, the overall error drops toward zero. Classical test theory is simpler than other methods (eg, item response theory, generalizability theory) and therefore satisfied the criterion of parsimony, which refers to choosing the simplest approach that meets one's measurement and evaluation needs [41].

The validity framework reflects consensus in the field of measurement and evaluation about what indicates the validity of a measure, particularly in domains such as education that focus on assessment. Specifically, validity refers to the extent to which evidence and theory support interpretations of the score for its proposed use [31]. The validity framework therefore proposes five ways in which a measure may or may not demonstrate validity: its content validity, its response process, its internal structure, its relationship to other variables, and the consequences of the measure [31,42]. Because our aim was to develop a new measure in an area with few metrics, our study directly addresses the first three of these five. We discuss how related and future research might inform the fourth and fifth ways of assessing validity.

Content Validity

Content validity (point 1 in the validity framework [31]) refers to how well items match the definition of a construct. To ensure content validity of items, in our original systematic review, we had used foundational literature [15,16,18,43-45]; held monthly or bimonthly consultations in person and by teleconference over the course of 2 years within our interdisciplinary group of experts, including patients, caregivers, health professionals, academic researchers, and other stakeholders; and consulted with 15 additional experts outside the research team [13]. Discussions over the years of the project centered on the items themselves as well as prioritization of items according to their relevance within our conceptual framework.

Response Process

Response process (point 2 in the validity framework [31]) refers to quality control when using a measure [42]. In our case, it is the extent to which analysts are able to accurately and consistently assign a value to each item in the measure. We had refined the response process for each item through an iterative process of data extraction and data validation. This included consultation with 15 external experts and four rounds of pilot data extraction and refinement of response processes across randomly selected sets of five articles each time (total: 20 articles). We had also confirmed the accuracy of the extracted data with the authors of the original articles included in the systematic review and found very low rates of error [13].

Internal Structure

Internal structure (point 3 in the validity framework [31]) addresses to what extent items in a measure are coherent among

themselves and conform to the construct on which the proposed score interpretations are based. In our case, good internal structure would indicate that although the items are distinct, they are all measuring the same overall construct. We would therefore be able to detect patterns reflecting this construct. Specifically, processes that are more user-centered would score higher, and processes that are less user-centered would score lower. To assess this, we first identified which prioritized items formed a positive definite matrix of tetrachoric correlations. Tetrachoric correlations are similar to correlations between continuous variables (eg, Pearson correlations) but instead calculate correlations between dichotomous (ie, yes/no, true/false) variables. A matrix can be thought of as something like a table of numbers. A matrix of correlations is a square matrix, meaning it has the same number of rows as columns, in which any given row or column of the matrix represents a vector made up of an item's correlations with each of the other items in the set. The diagonal of the matrix will contain values of 1 because those cells represent each item's correlation with itself. Positive definite matrices are matrices that are able to be inverted. For readers unfamiliar with matrix algebra, a useful analogy may be that inversion is to matrices as division is to numbers. Inversion is possible when the vectors (in our case, vectors of tetrachoric correlations between potential items in the measure) that make up the matrix are sufficiently independent of each other. Matrix inversion is required to conduct principal component analysis.

We identified the items to compose the set whose correlations would make up the matrix by first rank ordering possible items in the data set according to their priority in our conceptual framework, using the expertise of our interdisciplinary team (see the Patient Partnership section). We then built the matrix in a stepwise fashion, adding items until the matrix of correlations was no longer invertible. Then, based on classical item analysis in which we required discrimination indices >0.2 [46-48], we formed a group of items with an acceptable value of Kaiser's measure of sampling adequacy (>0.6 [49]), meaning that they share enough common variance to allow principal component analysis. We then conducted this analysis with Varimax rotation. Using the resultant scree plot and content expertise based on our conceptual framework, we identified components that explained sufficient variance in the data, retaining items with loadings over 0.4 on at least one factor. We also performed classical item analysis to assess the resultant psychometric properties of the items in the measure. Finally, we used confirmatory factor analysis with unweighted least squares estimation to test our hypothesis of the existence of a latent construct of user-centeredness explaining the variance in the three components. In other words, we tested whether or not our data suggested that the components we found in our analysis shared a common root.

Applying the Measure Within the Data Set

We applied the resulting measure within the data set to examine and compare scores for the two groups of projects within the original study: patient decision aids, which could have been developed in any way, and other personal health tools that specifically described their design and development method as user- or human-centered design. To explore potential changes in design and development methods over time, we plotted scores within the two groups according to the year of publication of the first paper published about each project. To provide further information about the distribution of scores within the data set used to develop the measure, we calculated percentile ranks of the scores within the data set, applying the definition of a percentile rank that, for example, being in the 97th percentile indicates that the score was higher than 96% of those tested [50].

We conducted analyses in SAS, version 9.4 (SAS Institute Inc) and in R, version 3.3.2 (The R Foundation).

Patient Partnership

Patients and other stakeholders participated in every aspect of the research for this project overall as members of the research team. For the development of the measure, patient and caregiver partners were most involved in the prioritization of items for analysis.

Availability of Data and Materials

Data used in this study are available via Scholars Portal Dataverse [51].

Results

Items Retained in the User-Centered Design 11-Item Measure (UCD-11)

Out of 19 identified potential variables, we retained 11 items in a three-factor structure explaining 68% of the variance in the data, which refers to the variance within the 19 variables. The Kaiser's measure of sampling accuracy was 0.68, which is considered acceptable [49]. Each item is binary and is scored as either present or absent. Table 1 and Figure 2 show the 11 retained items and factor structure. The Cronbach alpha for all 11 items was .72, indicating acceptable internal consistency [52].

Table 1. Final measure with factor loadings.

Items[a]	Explanations and examples	Factors		
		Preprototype involvement	Iterative responsiveness	Other expert involvement
1. Were potential end users (eg, patients, caregivers, family and friends, surrogates) involved in any steps to help understand users (eg, who they are, in what context might they use the tool) and their needs?	Such steps could include various forms of user research, including formal or informal needs assessment, focus groups, surveys, contextual inquiry, ethnographic observation of existing practices, literature review in which users were involved in appraising and interpreting existing literature, development of user groups, personas, user profiles, tasks, or scenarios, or other activities	0.82	—[b]	—
2. Were potential end users involved in any steps of designing, developing, and/or refining a prototype?	Such steps could include storyboarding, reviewing the draft design or content before starting to develop the tool, and designing, developing, or refining a prototype	0.83	—	—
3. Were potential end users involved in any steps intended to evaluate prototypes or a final version of the tool?	Such steps could include feasibility testing, usability testing with iterative prototypes, pilot testing, a randomized controlled trial of a final version of the tool, or other activities	—	0.78	—
4. Were potential end users asked their opinions of the tool in any way?	For example, they might be asked to voice their opinions in a focus group, interview, survey, or through other methods	—	0.80	—
5. Were potential end users observed using the tool in any way?	For example, they might be observed in a think-aloud study, cognitive interviews, through passive observation, logfiles, or other methods	—	0.71	—
6. Did the development process have 3 or more iterative cycles?	The definition of a cycle is that the team developed something and showed it to at least one person outside the team before making changes; each new cycle leads to a version of the tool that has been revised in some small or large way	—	0.64	—
7. Were changes between iterative cycles explicitly reported in any way?	For example, the team might have explicitly reported them in a peer-reviewed paper or in a technical report. In the case of rapid prototyping, such reporting could be, for example, a list of design decisions made and the rationale for the decisions	—	0.87	—
8. Were health professionals asked their opinion of the tool at any point?	Health professionals could be any relevant professionals, including physicians, nurses, allied health providers, etc. These professionals are not members of the research team. They provide care to people who are likely users of the tool. Asking for their opinion means simply asking for feedback, in contrast to, for example, observing their interaction with the tool or assessing the impact of the tool on health professionals' behavior	—	—	0.80
9. Were health professionals consulted before the first prototype was developed?	Consulting before the first prototype means consulting prior to developing anything. This may include a variety of consultation methods	0.49	—	0.75
10. Were health professionals consulted between initial and final prototypes?	Consulting between initial and final prototypes means some initial design of the tool was already created when consulting with health professionals	—	—	0.91
11. Was an expert panel involved?	An expert panel is typically an advisory panel composed of experts in areas relevant to the tool if such experts are not already present on the research team (eg, plain language experts, accessibility experts, designers, engineers, industrial designers, digital security experts, etc). These experts may be health professionals but not health professionals who would provide direct care to end users	—	—	0.56

[a]All items are scored as yes=1 and no=0. When assigning scores from written reports of projects, if an item is not reported as having been done, it is scored as not having been done. The total score on the User-Centered Design 11-item scale (UCD-11) is the number of yes answers and therefore ranges from 0 to 11.

[b]Factor loadings <0.40 are not shown. This is because loadings <0.40 indicate that the item does not contribute substantially to that factor.

Figure 2. Items and scoring of the User-Centered Design 11-item measure (UCD-11).

The preprototype involvement factor included 2 items: (1) whether prospective users (ie, patient, family, caregiver, or surrogate users) were involved in steps that help tool developers understand users, and (2) whether prospective users were involved in the steps of prototype development. The iterative responsiveness factor included 5 items: (3) whether prospective users were asked for their opinions; (4) whether they were observed using the tool; (5) whether they were involved in steps intended to evaluate the tool; (6) whether the development process had 3 or more iterative cycles; and (7) whether changes between iterative cycles were explicitly reported. The other expert involvement factor included 4 items: (8) whether health professionals were asked for their opinion; (9) whether health professionals were consulted before the first prototype was developed; (10) whether health professionals were consulted between initial and final prototypes; and (11) whether an expert panel of nonusers was involved. As shown in Table 1, each of the 11 items is formulated as a question that can be answered by “yes” or “no,” and is assumed to be “no” if the item is not reported. The score is the number of “yes” answers and therefore ranges from 0 to 11.

Items Not Retained in UCD-11

The 8 items not retained due to a lack of sufficient explanation of variance were whether or not: (1) the users involved were currently dealing with the health situation, (2) a formal patient organization was involved, (3) an advisory panel of users was involved, (4) there were users who were formal members of the

research team, (5) users were offered incentives or compensation of any kind for their involvement (eg, cash, gift cards, payment for parking), (6) people who were members of any vulnerable population were explicitly involved [53], (7) users were recruited using convenience sampling, and (8) users were recruited using methods that one might use to recruit from populations that may be harder to reach (eg, community centers, purposive sampling, snowball sampling).

Classical Test Theory and Confirmatory Factor Analysis Results

Classical item difficulty parameters ranged from 0.28 to 0.85 on a scale ranging from 0 to 1 and discrimination indices from 0.29 to 0.46, indicating good discriminating power [46-48]. This means that the items discriminate well between higher and lower overall scores on the measure. Confirmatory factor analysis demonstrated that a second-order model provided an acceptable to good fit [54] (standardized root mean residual=0.09; goodness of fit index=0.96; adjusted goodness of fit index=0.94; normed fit index=0.93), supporting our hypothesis of a latent construct of user-centeredness that explains the three factors. This means that UCD-11 provides a single score or a single number rather than multiple numbers, and may therefore be used as a unidimensional measure. Had we not observed a single latent construct, the measure would have always needed to be reported with scores for each factor.

Scores Within the Data Set

As expected when applying a measure to the data set used to develop it, scores within the data set were distributed across the full range of possible scores (ie, 0 to 11). The median score was 6 out of a possible 11 (IQR 3-8) across all 348 projects. Median scores were 5 out of a possible 11 (IQR 3-8) for the design and development of patient decision aids, and 7 out of a possible 11 (IQR 5-8) for other personal health tools in which the authors specifically described their design and development method as user- or human-centered design. The 95% CI of the difference in mean scores for patient decision aid projects compared to projects that described their approach as user- or human-centered design was (–1.5 to –0.3). Figure 3 shows scores over time within the two groups. There were no discernable time trends in UCD-11 scores.

Table 2 provides percentiles for each possible UCD-11 score within the data set of 348 projects.

Figure 3. User-Centered Design 11-Item scale (UCD-11) scores by year of publication of the first paper describing a project. UCD refers to other personal health tools explicitly naming user- or human-centered design as the guiding process. PtDA: patient decision aids.

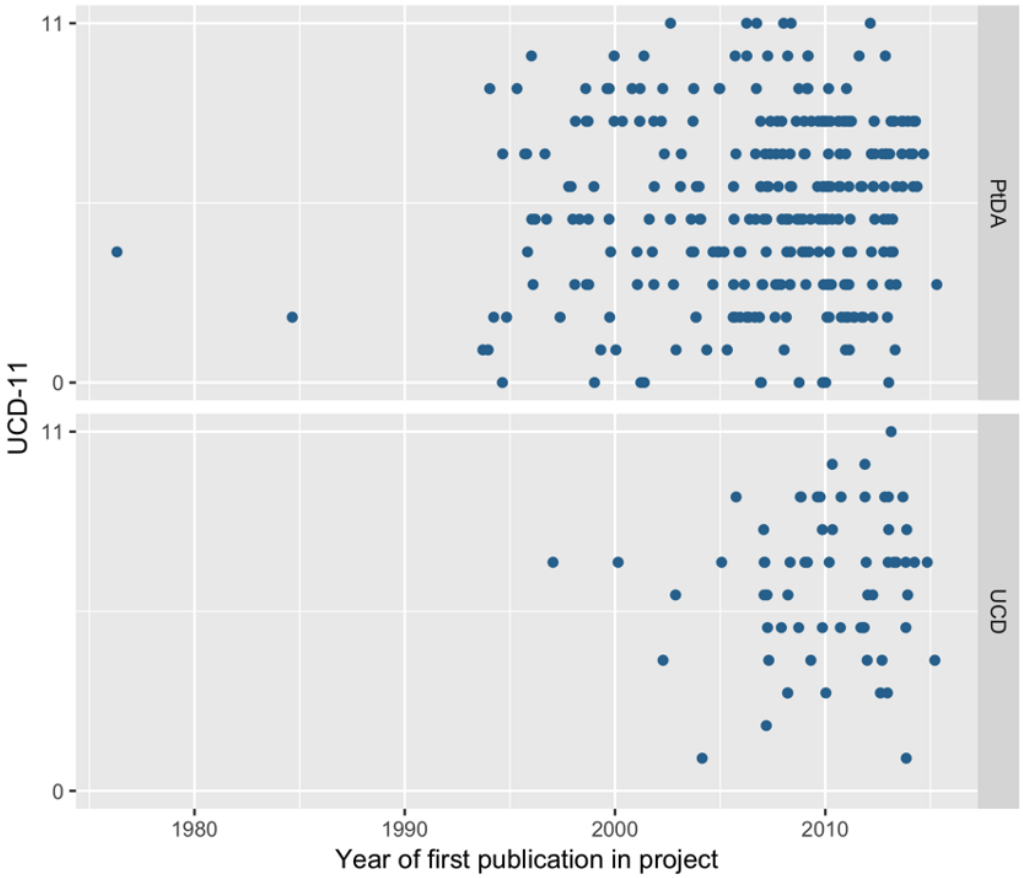


Table 2. Percentile ranks of the User-Centered Design 11-item scale (UCD-11) scores.

UCD-11 score	Percentile rank	Interpretation
0	0th	The score is not higher than any other scores in the data set.
1	4th	The score is higher than 3% of scores in the data set.
2	8th	The score is higher than 7% of scores in the data set.
3	17th	The score is higher than 16% of scores in the data set.
4	27th	The score is higher than 26% of scores in the data set.
5	36th	The score is higher than 35% of scores in the data set.
6	49th	The score is higher than 48% of scores in the data set.
7	61st	The score is higher than 60% of scores in the data set.
8	74th	The score is higher than 73% of scores in the data set.
9	87th	The score is higher than 86% of scores in the data set.
10	95th	The score is higher than 94% of scores in the data set.
11	99th	The score is higher than 98% of scores in the data set.

Discussion

Principal Results and Comparisons With Prior Work

Our study aimed to derive a measure of user-centeredness of the design and development processes for personal health tools. Applying a conceptual framework of user-centered design allowed us to identify indicators of this construct and develop an internally valid measure. This measure includes items that address the involvement of users and health professionals at every stage of a framework of user-centered design [14] as well as the importance of designing and developing tools in iterative cycles. Given the creative nature of design and development and a wide range of possible tools, the items are high-level assessments of whether or not particular aspects of involvement were present or absent, not assessments of the quality of each aspect.

To the best of our knowledge, ours is the first such validated measure for health applications. Other broadly applicable measures exist that assess the usability or ease of use of tools (eg, the System Usability Scale [55,56]). However, this measure assesses the quality of the resulting tool or system, not the process of arriving at the end product. Process measures do exist, for example, in software, consumer product development, and information systems [57-59].

Barki and Hartwick [54] developed measures centered around the design and development of information systems in professional contexts, with items reported by users. The items in their measures included “I was able to make changes to the formalized agreement of work to be done during system definition” and “I formally reviewed work done by Information Systems/Data Processing staff during implementation.” Users also indicated, for example, to what extent they felt the system was needed or relevant to them. Our measure has some items similar to the items in their user participation scale; however, in our measure, users themselves do not need to indicate whether or not a step occurred.

Kujala [58] offers a measure intended to assess the quality of system specifications after these have been developed. Items include “Customer or user requirements are completely defined” and “The correctness of the requirements is checked with real users,” assessed on a 4-point Likert scale, with responses ranging from “disagree” to “agree.” This measure assesses the quality of user research outputs, which should typically be generated early in a project. In contrast, our measure offers a means of measuring user involvement by the aspects of a design and development process that were or were not done during the entire process.

Subramanyam and colleagues [59] assessed user participation in software development using data collected from time sheets and surveys across 117 projects conducted over 4 years at a large manufacturing firm. Projects often consisted of developing manufacturing and supply chain software. They found that users reported higher satisfaction in projects developing new software when the demands on their time were lowest, whereas developers reported higher satisfaction when users’ time spent in the project was highest. Users in this case were employees in the firm, who presumably had other work-related tasks to do as well. Our measure differs from this approach in that we assess involvement in a variety of steps as well as other factors (eg, 3 or more iterative cycles) rather than the total time spent by users.

In summary, our measure aligns somewhat with work from other contexts to measure user-centeredness. The key difference between our measure and previous measures is that ours assesses the process of design and development rather than the quality of the end product, is specific to the context of health-related tools rather than that of information systems or more general contexts, and may be reported or assessed by anyone with sufficient knowledge of the design and development process rather than requiring reporting by users. This latter difference offers flexibility of administration and feasibility for assessing the design and development of completed projects. However, this also means that our measure does not capture the quality of involvement, neither from the perspectives of those involved nor in any sort of external way. Future research should compare

the relationship—or lack thereof—between whether or not specific steps occurred in a design and development process and users' perspectives on the quality of the design and development process. We also suggest that future research focused on the quality of the process might investigate how or whether including experts in design improves the design and development process and resulting tool. Previous research in tools designed for clinicians has shown that including design and human factors engineering experts generally increases the quality of the tools, and also that the extent of improvement varies considerably according to the individual expert [60].

In addition to the strengths of our study, the first external use of our measure, conducted through advance provision of the measure to colleagues, offered some additional promising indications of its validity, specifically with respect to the fourth and fifth items of the validity framework (relationship to other variables and consequences of the measure) that were not possible to assess in our study. Higgins and colleagues [61] conducted a systematic review of 26 electronic tools for managing children's pain. They aimed to investigate the characteristics of tools still available for patients and families to use versus those that were no longer in use. They found that higher UCD-11 scores were associated with the tools still being available for use after the grant and project had ended [61].

Although case reports suggest that involving users in the design and development of health-related tools can lead to more usable, accepted, or effective tools [62,63], and, as mentioned above, emerging evidence suggests that higher scores on our measure are associated with more sustained availability of tools [61], we lack definitive evidence about the extent to which increasing user-centeredness may improve tools. It may be that there is a point beyond which it is either not feasible or not a good use of limited time and resources to increase involvement. For these reasons, UCD-11 should be considered descriptive, not normative.

Limitations

Our study has two main limitations. First, our data came from published reports, not direct capture of design and development processes. Although we have reason to believe the data are of high quality given our rigorous data validation and low rates of error [13], data from a systematic review of this nature may not contain full details of design and development processes. We chose to use these data because we believed they might offer valuable insights across hundreds of projects. Another research team might choose to draft a list of items from scratch, seek to apply them to new design processes, and validate a measure that way, one project at a time. Second, because our largest data source came from reports of the design and development of patient decision aids, our findings may be overly influenced by practices in the field of shared decision making and patient decision aids. We believe that this focus is appropriate for increasing user-centeredness in the context of health care. Shared decision making has been noted as "the pinnacle of patient-centered care" [64] and patient-centered care has been defined as "care that is respectful of and responsive to individual patient preferences, needs, and values," such that, "patient values guide all clinical decisions" [65], a definition that aligns precisely with the goals of shared decision making [66]. However, it is possible that, because patient decision aids are intended to be used to complement consultation with a health professional, this focus in our data may have led to overemphasis on the role of health professionals in developing tools for use by people outside the health system.

Using UCD-11

Our goal in developing UCD-11 was to offer a straightforward, descriptive measure that can be used by teams as part of reporting their own processes or alternatively by researchers who may apply it to written reports of design and development processes. UCD-11 is intended as a complement to—not a replacement for—detailed descriptions of the design and development processes of personal health tools and is intended to be applied at the end of a project. As stated earlier, it is a descriptive, not normative, measure. Although Higgins and colleagues [61] offered evidence that higher UCD-11 scores are associated with positive implementation outcomes of a personal health tool, we do not have evidence that higher scores necessarily indicate higher-quality design and development processes.

Conclusions

Using a framework of user-centered design synthesized from foundational literature, we were able to derive UCD-11, an internally valid descriptive measure of the user-centeredness of the design and development processes of personal health tools. This measure offers a structured way to consider design and development methods (eg, co-design) when creating tools with and for patients and caregivers. Through measurement and reporting, this measure can help collect evidence about user involvement in order that future research might better specify how we can make the best possible use of the time and effort of all people involved in design and development. We hope this measure will help generate structured data toward this goal and help foster more creation of tools that are adapted to the people who will use them, rather than requiring people to adapt to the tools.

Acknowledgments

As with each of our team's papers, all members of the team were offered the opportunity to coauthor this publication. Not all members elected to accept the invitation, depending on their interest and other commitments. The authors thank team members Sholom Glouberman (patient team member), Jean Légaré (patient team member), Carrie A Levin (patient decision aid developer team member), Karli Lopez (caregiver team member), Victor M Montori (academic team member), and Kerri Sparling (patient team member), who all contributed to the broader project that led to the work presented here.

This study was funded by the Patient-Centered Outcomes Research Institute (PCORI): ME-1306-03174 (PI: HOW) and the Canadian Institutes of Health Research (CIHR): FDN-148246 (PI: HOW). PCORI and CIHR had no role in determining the study design, the plans for data collection or analysis, the decision to publish, nor the preparation of this manuscript. HOW is funded by a Tier 2 Canada Research Chair in Human-Centred Digital Health and received salary support during this work from Research Scholar Junior 1 and 2 Career Development Awards by the Fonds de Recherche du Québec—Santé (FRQS). AMCG received salary support during this work from a Research Scholar Junior 2 Career Development Award by the FRQS. NMI is funded by a Tier 2 Canada Research Chair in Implementation of Evidence-based Practice and received salary support during this work from a New Investigator Award by the CIHR as well as a New Investigator Award from the Department of Family and Community Medicine, University of Toronto. FL is funded by a Tier 1 Canada Research Chair in Shared Decision Making and Knowledge Translation. DS holds a University of Ottawa Research Chair in Knowledge Translation to Patients.

During the course of this project, Carrie A Levin (patient decision aid developer team member) received salary support as research director for the Informed Medical Decisions Foundation, the research division of Healthwise, Inc, a not-for-profit organization that creates products including patient decision aids.

Authors' Contributions

HOW, GV, and JSR were responsible for study conceptualization and methodology; JSR for validation; HOW, GV, and JSR for formal analysis; HOW, GV, SCD, HC, MD, AF, AMCG, LH, AH, NMI, FL, TP, DS, MET, RJV, and JSR for investigation; GV and TP for data curation; HOW and JSR for writing the original draft; HOW, GV, SCD, HC, MD, AF, AMCG, LH, AH, NMI, FL, TP, DS, MET, RJV, and JSR for writing, review, and editing; HOW and SCD for project administration; and HOW, SCD, HC, AF, AMCG, LH, AH, NMI, FL, DS, RJV, and JSR for funding acquisition.

Conflicts of Interest

No conflicts to declare.

References

1. Carrington P, Hurst A, Kane S. Wearables and Chairables: Inclusive Design of Mobile Input and Output Techniques for Power Wheelchair Users. New York, NY, USA: Association for Computing Machinery; 2014 Apr Presented at: Proceedings of the 14th SIGCHI Conference on Human Factors in Computing Systems; 2014; Toronto, ON, Canada p. 3103-3112. [doi: 10.1145/2556288.2557237]
2. Renda G, Jackson S, Kuys B, Whitfield TWA. The cutlery effect: do designed products for people with disabilities stigmatise them? Disabil Rehabil Assist Technol 2016 Nov;11(8):661-667. [doi: 10.3109/17483107.2015.1042077] [Medline: 25936430]
3. Heller S, White D, Lee E, Lawton J, Pollard D, Waugh N, et al. A cluster randomised trial, cost-effectiveness analysis and psychosocial evaluation of insulin pump therapy compared with multiple injections during flexible intensive insulin therapy for type 1 diabetes: the REPOSE Trial. Health Technol Assess 2017 Apr;21(20):1-278 [FREE Full text] [doi: 10.3310/hta21200] [Medline: 28440211]
4. Wallace EL, Lea J, Chaudhary NS, Griffin R, Hammelman E, Cohen J, et al. Home Dialysis Utilization Among Racial and Ethnic Minorities in the United States at the National, Regional, and State Level. Perit Dial Int 2017;37(1):21-29. [doi: 10.3747/pdi.2016.00025] [Medline: 27680759]
5. Shilkrot R, Huber J, Liu C, Maes P, Nanayakkara S. FingerReader: A Wearable Device to Support Text Reading on the Go. New York, NY, USA: Association of Computing Machinery; 2014 Apr Presented at: Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems (CHI EA '14); 2014; Toronto, ON, Canada. [doi: 10.1145/2559206.2581220]
6. Hawley MS, Cunningham SP, Green PD, Enderby P, Palmer R, Sehgal S, et al. A voice-input voice-output communication aid for people with severe speech impairment. IEEE Trans Neural Syst Rehabil Eng 2013 Jan;21(1):23-31. [doi: 10.1109/TNSRE.2012.2209678] [Medline: 22875259]
7. Markham SA, Levi BH, Green MJ, Schubart JR. Use of a Computer Program for Advance Care Planning with African American Participants. Journal of the National Medical Association 2015 Feb;107(1):26-32. [doi: 10.1016/S0027-9684(15)30006-7] [Medline: 27282525]
8. Hightow-Weidman LB, Muessig KE, Pike EC, LeGrand S, Baltierra N, Rucker AJ, et al. HealthMpowerment.org: Building Community Through a Mobile-Optimized, Online Health Promotion Intervention. Health Educ Behav 2015 Aug;42(4):493-499 [FREE Full text] [doi: 10.1177/1090198114562043] [Medline: 25588932]
9. Sridhar A, Chen A, Forbes E, Glik D. Mobile application for information on reversible contraception: a randomized controlled trial. Am J Obstet Gynecol 2015 Jun;212(6):774.e1-774.e7. [doi: 10.1016/j.ajog.2015.01.011] [Medline: 25582097]
10. Hartzler AL, Izard JP, Dalkin BL, Mikles SP, Gore JL. Design and feasibility of integrating personalized PRO dashboards into prostate cancer care. J Am Med Inform Assoc 2016 Jan;23(1):38-47 [FREE Full text] [doi: 10.1093/jamia/ocv101] [Medline: 26260247]

11. Sanchez-Morillo D, Fernandez-Granero M, Jiménez AL. Detecting COPD exacerbations early using daily telemonitoring of symptoms and k-means clustering: a pilot study. Med Biol Eng Comput 2015 May;53(5):441-451. [doi: 10.1007/s11517-015-1252-4] [Medline: 25725628]
12. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Apr 12;4:CD001431 [FREE Full text] [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
13. Vaisson G, Provencher T, Dugas M, Trottier ME, Chipenda Dansokho S, Colquhoun H, et al. User Involvement in the Design and Development of Patient Decision Aids and Other Personal Health Tools: A Systematic Review. Medical Decision Making (forthcoming) 2021. [doi: 10.1177/0272989X20984134]
14. Witteman HO, Dansokho SC, Colquhoun H, Coulter A, Dugas M, Fagerlin A, et al. User-centered design and the development of patient decision aids: protocol for a systematic review. Syst Rev 2015 Jan 26;4:11 [FREE Full text] [doi: 10.1186/2046-4053-4-11] [Medline: 25623074]
15. Gould JD, Lewis C. Designing for usability: key principles and what designers think. Communications of the ACM 1985;28(3):300-311. [doi: 10.1145/3166.3170]
16. Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. Switzerland: International Standardization Organization (ISO); Jul 2019:1-33.
17. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]
18. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications; 2004:445-456.
19. Garrett JJ. The elements of user experience, 2nd edition. Berkeley, CA: New Riders Publishing; 2011.
20. Tullis T, Albert W. Measuring the user experience. Waltham, MA: Morgan Kaufmann; 2010.
21. Goodman E, Kuniavsky M, Moed A. Observing the User Experience, 2nd edition. Waltham, MA: Morgan Kaufmann; 2012.
22. Iivari J, Iivari N. Varieties of user-centredness: An analysis of four systems development methods. Information Systems Journal 2011 Mar;21(2):125-153 [FREE Full text] [doi: 10.1111/j.1365-2575.2010.00351.x]
23. Strauss AL, Fagerhaugh S, Suczek B, Wiener C. The work of hospitalized patients. Social Science & Medicine 1982 Jan;16(9):977-986. [doi: 10.1016/0277-9536(82)90366-5] [Medline: 7112176]
24. Corbin J, Strauss A. Managing chronic illness at home: Three lines of work. Qual Sociol 1985;8(3):224-247. [doi: 10.1007/bf00989485]
25. Valdez RS, Holden RJ, Novak LL, Veinot TC. Transforming consumer health informatics through a patient work framework: connecting patients to context. J Am Med Inform Assoc 2015 Jan;22(1):2-10 [FREE Full text] [doi: 10.1136/amiajnl-2014-002826] [Medline: 25125685]
26. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. The invisible work of personal health information management among people with multiple chronic conditions: qualitative interview study among patients and providers. J Med Internet Res 2015 Jun 04;17(6):e137 [FREE Full text] [doi: 10.2196/jmir.4381] [Medline: 26043709]
27. Lochner K, Kawachi I, Kennedy BP. Social capital: a guide to its measurement. Health Place 1999 Dec;5(4):259-270. [doi: 10.1016/s1353-8292(99)00016-7] [Medline: 10984580]
28. Biggs J. What do inventories of students' learning processes really measure? A theoretical review and clarification. Br J Educ Psychol 1993 Feb;63 (Pt 1):3-19. [doi: 10.1111/j.2044-8279.1993.tb01038.x] [Medline: 8466833]
29. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec 9;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
30. Rubin HR, Pronovost P, Diette GB. The advantages and disadvantages of process-based measures of health care quality. Int J Qual Health Care 2001 Dec;13(6):469-474. [doi: 10.1093/intqhc/13.6.469] [Medline: 11769749]
31. American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational Psychological Testing. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association; 2014.
32. Guilford J. Psychometric Methods, 1st ed. New York and London: McGraw-Hill Book Company, Inc; 1936.
33. Gulliksen H. Theory of mental tests. New York: Wiley; 1950.
34. Lord F, Novick M, Birnbaum A. Statistical theories of mental test scores. Reading, MA: Addison-Wesley; 1968.
35. Magnusson D. Test Theory. Reading, MA: Addison-Wesley Pub Co; 1967.
36. Spearman C. The Proof and Measurement of Association between Two Things. Am J Psychol 1904 Jan;15(1):72-101. [doi: 10.2307/1412159]
37. Spearman C. Demonstration of Formulae for True Measurement of Correlation. Am J Psychol 1907 Apr;18(2):161-169. [doi: 10.2307/1412408]
38. Sébille V, Hardouin J, Le Néel T, Kubis G, Boyer F, Guillemin F, et al. Methodological issues regarding power of classical test theory (CTT) and item response theory (IRT)-based approaches for the comparison of patient-reported outcomes in

two groups of patients--a simulation study. BMC Med Res Methodol 2010 Mar 25;10:24 [FREE Full text] [doi: 10.1186/1471-2288-10-24] [Medline: 20338031]
39. Cappelleri JC, Jason Lundy J, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. Clin Ther 2014 May;36(5):648-662 [FREE Full text] [doi: 10.1016/j.clinthera.2014.04.006] [Medline: 24811753]
40. León-Del-Barco B, Mendo-Lázaro S, Felipe-Castaño E, Fajardo-Bullón F, Iglesias-Gallego D. Measuring Responsibility and Cooperation in Learning Teams in the University Setting: Validation of a Questionnaire. Front Psychol 2018;9:326 [FREE Full text] [doi: 10.3389/fpsyg.2018.00326] [Medline: 29593622]
41. De Champlain AF. A primer on classical test theory and item response theory for assessments in medical education. Med Educ 2010;44(1):109-117. [doi: 10.1111/j.1365-2923.2009.03425.x]
42. Downing SM. Validity: on meaningful interpretation of assessment data. Med Educ 2003 Sep;37(9):830-837. [doi: 10.1046/j.1365-2923.2003.01594.x] [Medline: 14506816]
43. Mao J, Vredenburg K, Smith PW, Carey T. The state of user-centered design practice. Commun ACM 2005 Mar;48(3):105-109. [doi: 10.1145/1047671.1047677]
44. Nielsen J. The usability engineering life cycle. Computer 1992 Mar;25(3):12-22. [doi: 10.1109/2.121503]
45. Norman D. The Design of Everyday Things. New York: Basic Books; 2002.
46. Nunnally J, Bernstein L. Psychometric theory. 3rd ed. New York: McGraw Hill; 1994.
47. Schmeiser CB, Welch CJ. Test development. Educational measurement 2006;4:307-353.
48. Everitt BS, Skrondal A. The Cambridge Dictionary of Statistics. Cambridge: Cambridge University Press; 2010.
49. Kaiser HF. An index of factorial simplicity. Psychometrika 1974 Mar;39(1):31-36. [doi: 10.1007/bf02291575]
50. Lang TA, Secic M. How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers. Philadelphia: American College of Physicians; 2006:20.
51. Data for Development and Validation of UCD-11: An 11-item Measure of User- and Human-Centered Design for Personal Health Tools. Scholars Portal Dataverse. URL: https://doi.org/10.5683/SP2/LJLUWQ [accessed 2021-02-22]
52. DeVellis R. Scale Development: Theory and Applications. 3rd ed. Thousand Oaks, CA: Sage Publications; 2011.
53. Dugas M, Trottier ME, Chipenda Dansokho S, Vaisson G, Provencher T, Colquhoun H, et al. Involving members of vulnerable populations in the development of patient decision aids: a mixed methods sequential explanatory study. BMC Med Inform Decis Mak 2017 Jan 19;17(1):12 [FREE Full text] [doi: 10.1186/s12911-016-0399-8] [Medline: 28103862]
54. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. Methods of Psychological Research Online 2003;8(8):23-74.
55. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
56. Bangor A, Kortum P, Miller J. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Journal of Usability Studies 2009 May;4(3):114-123.
57. Barki H, Hartwick J. Measuring User Participation, User Involvement, and User Attitude. MIS Quarterly 1994 Mar;18(1):59-82. [doi: 10.2307/249610]
58. Kujala S. Effective user involvement in product development by improving the analysis of user needs. Behaviour & Information Technology 2008 Nov;27(6):457-473. [doi: 10.1080/01449290601111051]
59. Subramanyam R, Weisstein F, Krishnan MS. User participation in software development projects. Commun. ACM 2010 Mar;53(3):137-141 [FREE Full text] [doi: 10.1145/1666420.1666455]
60. Kealey MR. Impact of Design Expertise and Methodologies on the Usability of Printed Education Materials Internet. TSpace. Toronto, ON: University of Toronto; 2015. URL: https://tspace.library.utoronto.ca/handle/1807/70839 [accessed 2020-07-31]
61. Higgins KS, Tutelman PR, Chambers CT, Witteman HO, Barwick M, Corkum P, et al. Availability of researcher-led eHealth tools for pain assessment and management: barriers, facilitators, costs, and design. PR9 2018 Sep;3(1):e686. [doi: 10.1097/pr9.0000000000000686]
62. Kilsdonk E, Peute LW, Riezebos RJ, Kremer LC, Jaspers MWM. From an expert-driven paper guideline to a user-centred decision support system: a usability comparison study. Artif Intell Med 2013 Sep;59(1):5-13. [doi: 10.1016/j.artmed.2013.04.004] [Medline: 23684240]
63. Wilkinson CR, De Angeli A. Applying user centred and participatory design approaches to commercial product development. Design Studies 2014 Nov;35(6):614-631. [doi: 10.1016/j.destud.2014.06.001]
64. Barry MJ, Edgman-Levitan S. Shared Decision Making — The Pinnacle of Patient-Centered Care. N Engl J Med 2012 Mar;366(9):780-781. [doi: 10.1056/nejmp1109283]
65. Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Journal For Healthcare Quality 2002;24(5):52. [doi: 10.1111/j.1945-1474.2002.tb00463.x]
66. Légaré F, Witteman HO. Shared decision making: examining key elements and barriers to adoption into routine clinical practice. Health Aff (Millwood) 2013 Feb;32(2):276-284. [doi: 10.1377/hlthaff.2012.1078] [Medline: 23381520]

Abbreviations

mHealth: mobile health
UCD-11: User-Centered Design 11-item scale

Edited by G Eysenbach; submitted 13.06.19; peer-reviewed by I Holeman, L van Velsen, D Sakaguchi-Tang; comments to author 03.10.19; revised version received 27.08.20; accepted 03.10.20; published 16.03.21.

Please cite as:
Witteman HO, Vaisson G, Provencher T, Chipenda Dansokho S, Colquhoun H, Dugas M, Fagerlin A, Giguere AMC, Haslett L, Hoffman A, Ivers NM, Légaré F, Trottier ME, Stacey D, Volk RJ, Renaud JS
An 11-Item Measure of User- and Human-Centered Design for Personal Health Tools (UCD-11): Development and Validation
J Med Internet Res 2021;23(3):e15032
URL: https://www.jmir.org/2021/3/e15032
doi:10.2196/15032
PMID:33724194

Original Paper

Adaptation of Extended Reality Smart Glasses for Core Nursing Skill Training Among Undergraduate Nursing Students: Usability and Feasibility Study

Sun Kyung Kim[1,2*], PhD; Youngho Lee[3*], PhD; Hyoseok Yoon[4*], PhD; Jongmyung Choi[3], PhD

[1]Department of Nursing, Mokpo National University, Muan, Republic of Korea

[2]Department of Biomedicine, Health & Life Convergence Sciences, BK21 Four, Mokpo National University, Muan, Republic of Korea

[3]Department of Computer Engineering, Mokpo National University, Muan, Republic of Korea

[4]Division of Computer Engineering, Hanshin University, Osan, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Sun Kyung Kim, PhD
Department of Biomedicine
Health & Life Convergence Sciences, BK21 Four
Mokpo National University
1666, Yeongsan-ro, Cheonggye-myeon
Muan-gun, Jeollanam-do
Muan
Republic of Korea
Phone: 82 614506292
Email: skkim@mokpo.ac.kr

Abstract

Background: Skill training in nursing education has been highly dependent on self-training because of Korea's high student-faculty ratio. Students tend to have a passive attitude in self-practice, and it is hard to expect effective learning outcomes with traditional checklist-dependent self-practice. Smart glasses have a high potential to assist nursing students with timely information, and a hands-free device does not interrupt performance.

Objective: This study aimed to develop a smart glass–based nursing skill training program and evaluate its usability and feasibility for the implementation of self-practice.

Methods: We conducted a usability and feasibility study with 30 undergraduate nursing students during a 2-hour open lab for self-practice of core nursing skills, wearing smart glasses for visualized guidance. The usability test was conducted using a 16-item self-reporting questionnaire and 7 open-ended questions. Learning satisfaction was assessed using a 7-item questionnaire. The number of practice sessions was recorded, and perceived competency in core nursing skills was measured before and after the intervention. At the final evaluation, performance accuracy and time consumed for completion were recorded.

Results: Smart glass–assisted self-practice of nursing skills was perceived as helpful, convenient, and interesting. Participants reported improved recollection of sequences of skills, and perceived competency was significantly improved. Several issues were raised by participants regarding smart glasses, including small screen size, touch sensors, fogged lenses with masks, heaviness, and heat after a period of time.

Conclusions: Smart glasses have the potential to assist self-practice, providing timely information at students' own paces. Having both hands free from holding a device, participants reported the convenience of learning as they could practice and view the information simultaneously. Further revision correcting reported issues would improve the applicability of smart glasses in other areas of nursing education.

(J Med Internet Res 2021;23(3):e24313) doi:10.2196/24313

KEYWORDS

nursing education; skill training; self-practice; smart glass; usability; feasibility

Introduction

Advancements in life science and biotechnology have transformed the hospital environment, and the need for qualified health professionals has never been higher [1]. In most countries, nurses comprise the largest proportion of the hospital workforce; thus, nurse staffing with a sufficient number of skilled nurses is essential for quality care [2]. It is the responsibility of nursing schools at the undergraduate level to ensure quality care and patient safety with a well-organized curriculum. Practical training is as necessary as theoretical nursing education, and the importance of nursing skill training has been well recognized [3].

Acquisition of mastery in clinical nursing skills not only improves the overall quality of patient care, but also leads to a successful and confident nursing career [4]. Given the growing need for better assurance of practical ability, the Korean Accreditation Board of Nursing Education (KABONE) [5] identified 20 core nursing skills in which nursing students are expected to attain a good level of performance in their accreditation process [6]. Although performance exams have been widely implemented in other licensure examinations and for nurse licensure in other countries such as Canada [7], nursing education in Korea has fully relied on self-practice.

Regardless of the importance of skill training, challenges exist for running educational training programs in Korea. Schools lack the ability to accommodate necessary training because of high student-teacher ratios, so self-practice has been introduced as an alternative method for nursing skill training [8,9]. Considering the insufficient coaching and supervision, educational strategies are needed to improve the effectiveness and efficacy of this self-practice.

At present, students use a written checklist for self-practice, which provides text-based descriptions for each step. However, using these checklists without proper instruction from lecturers means that there is a high risk that students will misconduct their self-practice and repeat wrong performances. Prior studies have indicated that text alone is limited in delivering messages clearly when it contains complex issues [10,11]. The use of an image, which may be worth a thousand words [12], can make complex processes visible, which could effectively reduce the cognitive load involved in acquisition of skills [13]. Knowing that precise step-by-step implementation is essential, both comprehensive understanding and perfect memorization of each step would ensure excellent skill performance.

Visualization can be an effective solution, and its value for better learning engagement and active learning has been used recognized in education [14]. A benefit of visual representation is that a complicated process can be memorized more easily with graphically illustrated essential concepts. Currently, there has been a growing interest in using extended reality (XR) technology for training health care professionals [15]. With improvements in technology involving wearable devices, such as smart glasses, XR has been applied in many health care training programs [16-19]. The findings of these studies showed promising outcomes as effective alternatives to traditional educational programs. XR technology allows for new learning experiences via superimposition of holographic visualization on what users see in the real world.

Previous studies found passive attitudes among students participating in self-training programs, leading to a lack of competency in future nursing practice [10,20]. Furthermore, knowing that the training was insufficient, students lacked competency after completion of this unattended training, causing them anxiety and stress in nursing practice. XR technologies could be a solution, effectively assisting students' self-training so that students are more likely to perceive that self-training is well structured and of high quality. In addition, the burden on faculty members to provide individual guidance can be alleviated because instant correction, where students reflect on timely information provided by smart glasses, is possible. Effective delivery of visualized education materials via an XR device could potentiate the learning experience without excessive consumption of educational manpower for supervision.

Along with visualization, timely information facilitates skill acquisition and completion. It is necessary to provide the experience of performing a true-to-life working process [21]. In addition, interacting with advanced technology could facilitate students' motivation for self-practice. Previous studies showed that higher levels of attention and better learning engagement were achieved when implementing XR in education [22-24]. Using smart glasses enhances users' engagement in performance [25]. Smart glasses improve the efficiency of practice, helping students master each skill with timely information without compromising performance. For complex skills, favorable consequences are expected to be higher, allowing students to experience a sense of accomplishment, completing exercises in a perfect manner, which would lead to improved competency in core nursing skills.

Smart glasses using augmented reality (AR) have previously been applied to support nursing care activities (eg, wound care management, mass casualty triage classification, and central line placement) [26]. These studies mostly focused on ease of obtaining knowledge and advanced features assisting the smart glasses' performance. The positive implications of using smart glasses to assist in nursing activities were assured. The purpose of our study was to test the feasibility and usability of implementing a core nursing skill training program that combined visualization and XR technology for undergraduate nursing students. We hypothesized that a smart glass–based nursing skill training program would not only assist practice but also induce active engagement of students into self-training.

Methods

Design of Graphical Images for Screens

We developed an XR image guide training program for 2 core nursing skills, specifically, blood transfusion and intradermal injection administration. Of the 20 core nursing skills listed, these 2 skills were randomly chosen from those ranked high in difficulty level, classified by KABONE according to the procedures' complexity. The numerous steps of these skills were split into several graphical images to be displayed on smart

glasses. Each graphical image transposed to the smart glasses paralleled the text information in the original checklists. The contents of the XR image guide training program are shown in Table 1. The contents were developed and revised several times, considering conciseness and adequacy, which involved expert review by a team of 3 nursing faculty members (2 experts in fundamental nursing and 1 in nursing informatics) and user evaluation by nursing students. Students who participated in this user evaluation were asked whether the meaning was well delivered and could be recognized at a glance without misunderstanding the graphical image. The image was drawn as concisely as possible because of the small capacity of the screen within smart glasses. The conformity between the final version of the graphical image and text information in the checklists was reviewed by 2 professors who had more than 5 years of teaching experience in the fundamental nursing curriculum.

Table 1. Description of core nursing skills for smart glass–based self-practice.

Item	Steps, n	Core task	Necessary equipment and supplies
Blood transfusion	23	1. Preparing equipment and supplies. 2. Connecting blood product and injecting at right rate. 3. Noticing and warning for possible side effects. 4. Recording the nursing practice and patients' conditions.	Number of items: 18 Manikin, blood product, alcohol swab, 3-way stopcock, gloves, IV[a] pole, tray, watch with a second hand, stethoscope, sphygmomanometer, thermometer, kidney basin, recording paper, letter of consent, document needs sign, containers for general medical and damageable waste, and hand sanitizer.
Administration of intradermal injection	27	1. Preparing equipment and supplies. 2. Making diluted solution for AST.[b] 3. Administering of intradermal injection 4. Reading the results of skin test.	Number of items: 12 Prescription, two 1-ml & 5-ml syringes, alcohol swab, manikin, vial, ampoule of normal saline, tray, recording paper, containers for general medical and damageable waste, and hand sanitizer.

[a]IV: intravenous.

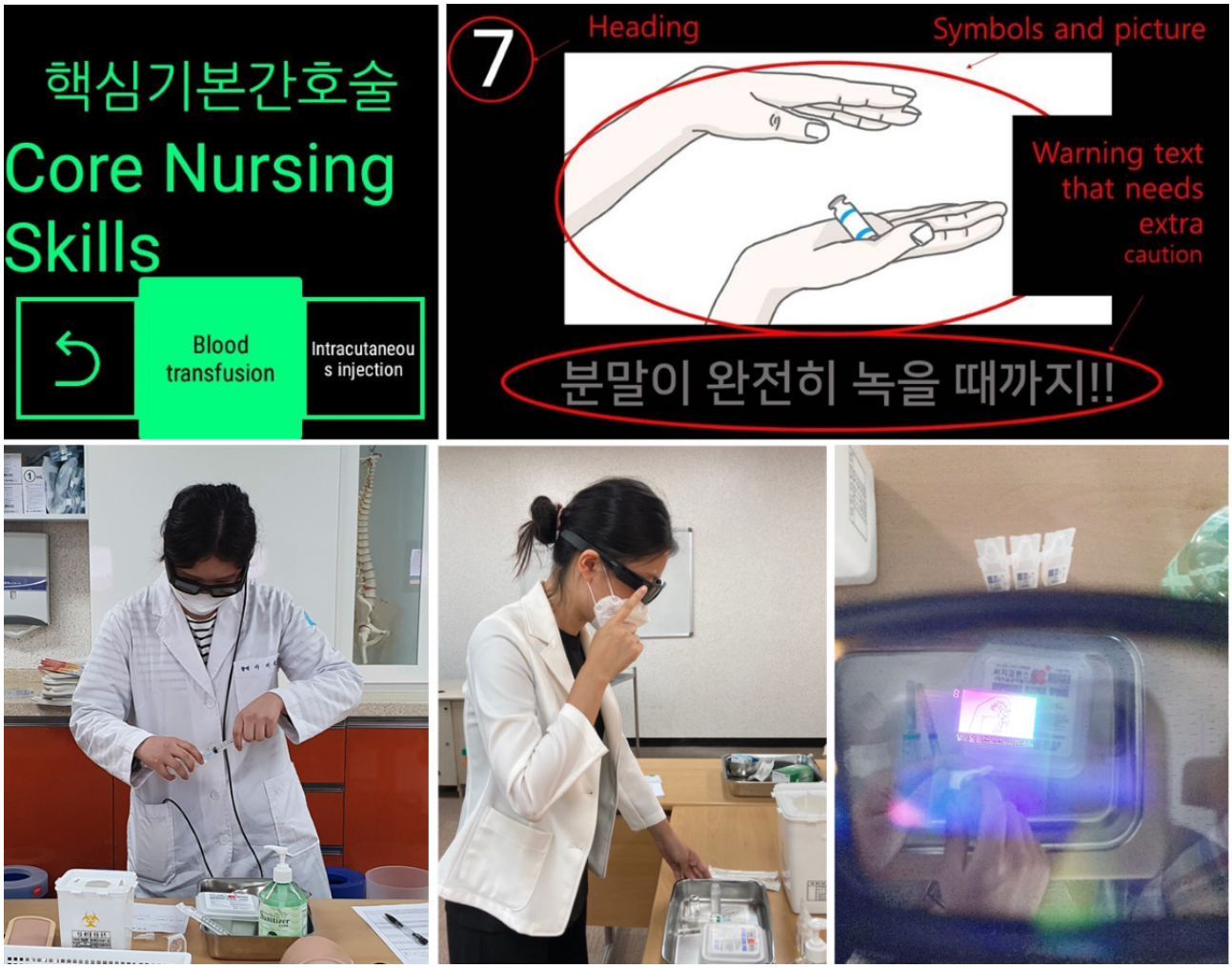
[b]AST: antibiotic skin test.

Preparation for User Study With a Smart Glass–Based Self-Training Program

A group of 5 students was given 2 hours of self-training and shared 2 Vuzix smart glasses. Students were encouraged to use the smart glasses at least once, with scheduled turns for the first use. According to KABONE, the estimated time to complete each skill was 10 minutes [5]. Before the beginning of their practice, brief instructions regarding how to operate the smart glasses (eg, content of the screen, how to find an item in the menu, turning to the next image) were provided, and individual students were given opportunities to wear the glasses and practice for about 5 minutes. After 2 hours of self-training, a performance test regarding accuracy and proficiency was conducted by an educator who had more than 5 years of experience in nursing education. During the test, the time for performance completion was measured by a research assistant.

Each image slide consists of 2 or 3 zones (Figure 1): (1) heading with sequence of the present step, (2) symbols and pictures representing the required performance in the present step, and (3) warning text that a step needs extra caution (when necessary).

Figure 1. Composition and contents of smart glass displays.



The graphical images appeared in the order of actions following the sequence of the original text-based checklists. The 2 core skills of blood transfusion and intradermal injection administration were adapted to 23 and 27 screens, respectively (Figures S1 and S2 in Multimedia Appendix 1).

Implementation of Our Smart Glass Application and Interface Design

The Vuzix Blade has a display only in the right eye, the display size is 480 × 853, and the shape of the display is a square. Vuzix's appearance is similar to that of ordinary glasses, and it supports voice recognition and touchpads. It supports Bluetooth and Wi-Fi networks and has a camera attached to take photos and videos or engage in remote collaboration. It is a stand-alone device that weighs 93.6 g and does not require additional equipment. It has a screen saver, so users can use it like a transparent glass in normal times and turn on the display when they need information. The Vuzix Blade currently runs Android 5.1, which supports application programming interface (API) 22 for the target API. Developers can develop the software using Java or Kotlin using Android Studio. Differences from general Android programming include the voice recognition API, touch interface API, and heads-up display API for graphical user interface [27].

We set up our device as an always-on display. In general, smart glass displays are on-demand displays that turn off the screen after a certain period to save power. When information is only available for a short time upon request, job performance declines because of psychological pressure [28]. If the students had to touch the touchpad or call a voice command every time they requested information, it would waste their time, and they could become exhausted by simple repetitive tasks. Thus, we turned off the screen saver and kept the display on while the students practiced.

The Vuzix Blade allows user interface elements to be navigated with simple left/right/up/down navigation. The menu is expressed in a square shape at the bottom of the screen. Students can select a submenu by swiping and tapping the touchpad. When a student chooses a submenu, the corresponding image is displayed. We implemented a simple input method to reduce malfunctioning when students use the touchpad. When selecting a task in the list, they are only allowed to use the left/right swipe and one-finger tap. When flipping the slide, only left/right swipes were permitted. It was originally set up to swipe when moving to the upper menu, but we assigned a two-finger tap for moving to the upper menu.

Usability Test

Seventeen items were used for the quantitative usability test. Items were developed based on previous studies in which

relevant items were selected and revised to be aligned with the purpose and methodology of this study. The study participants reported perceived usefulness items and ease of use items on 5-point scales, from strongly disagree (1 point) to strongly agree (5 points).

Learning Satisfaction

Level of satisfaction was assessed using 7 questions rated on a scale of 1 (strongly disagree) to 10 (strongly agree). Developed by Ji and Chung [29], questions were modified to fit the nursing education program best. With a maximum score of 70, a higher score indicates greater satisfaction with the education program.

Nursing Competency

Levels of perceived competency on 2 core nursing skills (administration of intradermal injection and transfusion) were assessed using a 10-point Likert scale. Developed by Han, Cho, and Won [30], a higher score indicates a greater level of competency for each skill.

Observation Data

During the 2-hour self-practice program, the overall number of practice attempts and number of practice attempts wearing smart glasses were observed and recorded by a research assistant. Developed by KABONE, a standardized checklist was used to measure the performance of 2 nursing skills (administration of intradermal injection and transfusion). The checklist consists of procedures of each skill, from preparing materials to writing nursing records after completion of skills. Scoring ranged from 0 to 100; a higher score indicates more accurate and precise performance without mistakes or omitted steps. At the nursing skill performance examination, individuals' time spent on performance completion was measured and recorded by a research assistant.

Qualitative Responses

Seven qualitative questionnaires were used to obtain comprehensive and detailed information about students' experiences using smart glasses for core skill nursing training. The questionnaire included the following: (1) How did you find the smart glass–based training in general? (2) Was this program easy to use? Did you need additional instructions? (3) Was there any content causing confusion or difficulties? (4) Did you experience any difficulties while operating smart glasses? (5) Do you think it will be helpful for your future clinical practice? If so, how? (6) Would you make any recommendations that are needed to improve this training program? (7) If you have any other comments regarding this smart glass–based training program, feel free to add them.

Ethical Considerations

The application of smart glass–based core nursing training for undergraduate nursing students was approved by the institutional review board (IRB no. MNUIRB-200326-BM-004-02) at a national university in Korea. Informed consent was obtained prior to obtaining the pretest data, and participants were told that they could stop participating anytime they wanted.

Statistical Analysis

Quantitative statistical data analysis was conducted using SPSS (version 25.0; IBM Corp). Mean, SD, frequency, and percentage were calculated for the demographic data, observation data, usability, and learning satisfaction survey. Paired 2-tailed *t* tests were used to compare outcome measures preintervention and postintervention, and an independent *t* test was conducted to identify between-groups differences. Pearson correlation analysis was conducted to determine the association between variables. For the qualitative data, all responses were reviewed and coded to identify common themes that were frequently reported.

Results

Quantitative Findings

The mean age of the study participants was 22.70 years, and 63% (19/30) were female. Approximately two-thirds of participants (22/30, 73%) reported possessing a moderate level of competency in core nursing skills, and approximately 1 in 10 (3/28, 11%) had previous experience with AR (Table 2).

Table 2. Demographic characteristics of study participants (N=30).

Characteristics	Value
Age (years), mean (SD)	22.70 (1.39)
Gender, n (%)	
Male	11 (37)
Female	19 (63)
Academic grade, n (%)	
Good	4 (13)
Fair	18 (60)
Poor	8 (27)
Satisfaction with clinical placement, n (%)	
High	21 (70)
Moderate	9 (30)
Low	0 (0)
Core nursing skill competency, n (%)	
Good	8 (27)
Fair	22 (73)
Poor	0 (0)
Previous experience with augmented reality, n (%)	
Yes	3 (11)
No	25 (89)

Participants were given 2 hours of open lab for self-practice of 2 core nursing skills. The number of practice attempts varied between 5 and 9. Full usage of the smart glasses during the self-practice open lab was observed and recorded. Participants used smart glasses in their practice as little as 2 and as many as 6 times (Table 3).

Table 3. Number of practice attempts and smart glass use.

Categories	Range	Mean (SD)
Practice attempts (total), n		
Blood transfusion	2-6	3.30 (0.952)
Administration of intradermal injection	2-5	3.73 (0.944)
Total	5-9	7.03 (1.25)
Practice attempts wearing smart glasses, n		
Blood transfusion	0-3	1.83 (0.747)
Administration of intradermal injection	0-3	1.70 (0.651)
Total	2-6	3.53 (0.973)

Regarding self-reported usability of the smart glass–based self-practice program, the highest score was obtained for question 11 (perceived interest) with a mean of 9.50 (SD 0.86). Question 5 (screen resolution) scored the lowest with a mean of 7.20 (SD 2.02). The degree of difficulties experienced with devices was rated with a mean of 3.83 (SD 2.73) (Table 4).

Table 4. Results of 16-item usability test (N=30).

Item	Range	Usability, mean (SD)
Ease of use		
1. How convenient do you think the smart glass–based core nursing education program is?	3-10	8.10 (1.58)
2. Was the initial education regarding the device and usage appropriate?	5-10	8.77 (1.46)
3. Was the text information presented on the screen easy to read?	3-10	7.27 (2.26)
4. Was the picture information presented on the screen clearly understood?	4-10	8.17 (1.90)
5. Was the resolution of the screen good?	4-10	7.20 (2.02)
6. Did you have any difficulties because of errors that occurred during the performance?	1-10	3.83 (2.73)
7. Was the progression speed adequate?	5-10	8.63 (1.22)
8. Was the location of the information on the smart glass appropriate? Consistent? Easy to see?	3-10	8.53 (1.85)
9. Was it convenient to operate the smart glass?	6-10	8.40 (1.48)
Usefulness		
10. Did the pictures and text information shown help you perform core basic nursing skills?	6-10	9.07 (1.05)
11. Was this type of educational program interesting?	7-10	9.50 (0.86)
12. Did you expect better scores using the smart glass training program?	6-10	8.90 (1.21)
13. Did you have a better understanding of core basic nursing techniques using augmented reality?	5-10	8.77 (1.48)
14. Would you recommend the smart glass–based core nursing education program to other friends?	3-10	8.70 (2.00)
15. Do you think smart glass core nursing education will be useful in clinical practice in the future?	4-10	8.77 (1.61)
16. Are you willing to use a smart glass for other core nursing skills in the future?	5-10	8.87 (1.57)

The overall score for learning satisfaction was 9.00 (SD 0.72). The participants gave the highest rating to “It was an interesting learning experience” (mean 9.60, SD 0.68) and the lowest rating to “It was more effective than lecturer-based education” (mean 7.43, SD 1.81) (Table 5).

Table 5. Results of 7-item learning satisfaction score (N=30).

Item	Range	Mean (SD)
1. It was an interesting learning experience.	8-10	9.60 (0.68)
2. Educational goals of this program were well-achieved.	7-10	9.30 (0.88)
3. It was a meaningful learning experience.	8-10	9.47 (0.73)
4. It was more effective than lecturer-based education.	3-10	7.43 (1.81)
5. I actively engaged in learning.	7-10	9.27 (0.91)
6. I felt satisfied with the educational program overall.	7-10	9.23 (0.90)
7. I hope to use this educational program for other subjects.	3-10	8.73 (1.46)
Learning satisfaction score (total)	N/A[a]	9.00 (0.72)

[a]N/A: not applicable.

Paired 2-tailed *t* tests were used to analyze differences between preintervention and postintervention competency of students on 2 nursing skills. Statistically significant improvement was achieved in both the skills after intervention ($P<.001$) as mean scores increased from 7.23 (SD 1.17) to 8.90 (SD 0.85) and 6.17 (SD 1.64) to 8.50 (SD 0.97), respectively (Table 6).

Table 6. Perceived competency in nursing skills before and after smart glass–based self-practice (N=30).

Item, time	Mean (SD)	*t* test (*df*)	*P* value
Administration of intradermal injection		−6.53 (29)	<.001
Pre	7.23 (1.17)		
Post	8.90 (0.85)		
Blood transfusion		−7.00 (29)	<.001
Pre	6.17 (1.64)		
Post	8.50 (0.97)		

The Pearson correlation test was conducted and identified statistically significant negative correlations between the number of practice attempts wearing smart glasses and performance time (r=−0.666, P<.001). There was a statistically significant positive correlation between the number of practice attempts wearing smart glasses and learning satisfaction (r=0.404, P=.03) (Table S1 in Multimedia Appendix 1). An independent t test revealed statistically significant gender differences for usability scores (ease of use, P=.049) and previous experience with devices (ease of use, P=.02; usefulness, P=.002) (Table S2 in Multimedia Appendix 1).

Qualitative Findings

Overall Experience of the Smart Glass–Based Skill Training Program

In general, the majority of students found smart glass–based skill training interesting (13/30, 43%) and convenient (8/30, 27%). About one in three participants (8/30, 27%) did not find significant benefits of using smart glasses for self-practice, and one participant reported, “It was interesting at first, but previous text based learning fits better for me.” Resistance to learning new technology was revealed, with participants saying, “I think I had to make more effort to learn about the devices.”

With regard to smart glasses, a large number of participants reported some degree of discomfort. There were touch sensor–related issues (9/30, 30%), specifically, “The touch sensor was too sensitive.” Others complained about the smart glass screen, and participants said, “Small sized text and low resolution caused eyestrain.”

Generally, participants responded that this smart glass–based self-training has educational benefits. Some participants found increased engagement in learning with new technology, saying “I was fascinated by the smart glasses and practice became more interesting using it.” Participants responded to the effectiveness of visualized information, improving their memory of educational content (n=7). Moreover, timely provision of information was found to have significant positive benefits (n=8): “Assisted by timely provided information, accurate and seamless practice was ensured using smart glasses.”

Perceived Easiness

Overall, the participants considered smart glass–based training to be very intuitive. Two out of three (20/30, 67%) reported immediate adaptation to smart glasses, with one participant saying “It was quite straightforward, I figured out how it works right away.” One-third of participants expressed the need for additional instructions, with one saying “I got confused with device, especially the touch pad on the glasses did not make sense for me.”

Recommendations

Several participants provided feedback regarding areas that needed further improvement. In terms of the smart glasses, participants raised the following issues: (1) touch sensor not working properly while wearing latex gloves, (2) glasses easily fogging while wearing a mask, (3) pain of double-layering glasses for people with poor eyesight, and (4) discomfort due to heavy weight and heat after a period of time (about 15 minutes). Regarding the training program content, participants reported tediousness of the simple text and image, saying “I expected something more entertaining like games.”

Discussion

Principal Findings

We explored the perceived usability and feasibility of smart glass–based self-practice among undergraduate nursing students. In general, the findings indicate that the participants had a greater degree of interest in this new device. Although some participants showed resistance to learning about the device, most students were pleased with having new educational methods to assist in their self-practice. This is closely linked to the characteristics of the study population. Recent advances in computing technology have transformed education, and the current generation is accustomed to this continuous change [26].

The findings of this study revealed the positive effects of smart glasses on engaging students in self-practice. Like self-practice, where an active learning attitude is essential, smart glasses could certainly provide a learner-centered education platform, allowing learning at an individual’s own pace without restrictions of time and supervisors [31]. This learner-centered approach enables students to make the best use of learning materials. As a matter of fact, participants who showed high levels of learning satisfaction used the smart glasses more frequently. This indicates that active participation is closely linked to the attractiveness of education strategies, which would eventually lead to learning satisfaction. Knowing that implementation of smart glasses induced students’ interest in skill practice, thoughtful consideration is required for nursing faculty members, who should make efforts to identify applicable areas where this technology can be used.

The findings of this study indicate greater improvement in perceived competency of performing blood transfusions. One possible explanation is variability in individual skills. Blood transfusion requires not only skillful performance but also complex procedures that involve multiple confirmation process (eg, physicians' orders, blood continuants, lab examination results, monitoring patients for side effects) [32] that frequently cause low levels of competency among undergraduate nursing students. Thus, by following guidance provided via smart glasses, students may find the tasks easier to complete. In addition, students would be better equipped with more complete knowledge regarding the step-by-step processes of blood transfusion, leading to increased competency.

In terms of number of uses, there was a positive correlation with learning satisfaction and a negative correlation with time consumed on performance completion. In addition, participants responded to educational benefits, as a timely graphical image assisted in improving their memory of the correct sequences. This suggests that the use of smart glasses has great potential to boost students' abilities for task completion. This is in line with previous studies on AR smart glasses, which assist health professionals in enabling simultaneous performance of multiple tasks [28,33]. One possible explanation is better cognitive performance due to greater efficacy in information confirmation provided via the glass screen and easy recollection of graphic-based guidance. Similarly, fast-tracking (shortening) the link between knowledge acquisition and actual practice or even tightly coupling them concurrently could expand individuals' existing abilities. Although usage frequency is closely related to students' preferences for this kind of learning, cumulative evidence would easily induce active engagement.

Several responses involved feedback that needs further consideration for the use of smart glasses to practice skills in nursing education, including low resolution, lack of visibility due to small-sized text, light smudging, perceived heaviness, severe condensation when wearing masks, and pain and discomfort for users wearing eyeglasses. Smart glasses have been considered for clinical implementation such as medical remote collaboration; thus, this is crucial information for future studies. Tasks are (1) determining the optimal size of text for users with poor eyesight, (2) identifying colors that cause light smudging, and (3) comparing and selecting smart glasses that are less likely to cause these issues. For example, the recently developed Google Glass may overcome some of these drawbacks with its advanced display, customizable hard case (lighter version without lenses or version with a thicker and solid frame), and lightweight form factor [34].

Findings of learning satisfaction outcomes revealed relatively lower scores on the item: "It was more effective than lecturer-based education." This indicates a limitation of the smart glass–based self-practice program. Although students' practice was assisted by the smart glasses, it could not sufficiently replace lecturer-based education. This implies the need for additional strategies to meet the educational needs of nursing skill training. Integration of prior educational strategies could effectively reinforce the current version of smart glass–based education.

The diverse features of smart glasses would more effectively replace previous strategies used for self-practice in nursing education. First, self-practice with video recording with self-feedback or peer feedback, the effectiveness of which was well established in a previous review study [35], could well be administered using smart glasses in a simpler and more convenient manner. Second, a demonstration from the lecturer can also be delivered via a smart glass guidance system. Inserting recorded videos or a series of photos in a GIF for demonstration could well guide students' acquisition of nursing skills. Lastly, consideration for aligning procedural steps with multimodal feedback, such as notification timer, sound, and vibration effect, is needed. There is evidence that gamification contributes to teaching and learning in nursing education [36], and these features have a high potential to immerse learners in self-training.

We conducted further statistical analysis (Tables S1 and S2 in Multimedia Appendix 1) to identify the potential influence of various user characteristics such as age, gender, and previous experience with devices. Interestingly, female participants and participants with no previous experience with AR devices reported better usability of the current smart glass–based training program. Previously, males were believed to be more willing to use and adapt more quickly to new technologies [37], while other studies observed no gender differences [38,39]. The findings of this study partially align with those of a recent study by Drin, Alamaki, and Soumala [40], which reported greater interest among females toward new technology. Novelty effects might be related to the lower usability scores of participants with previous experience; this may be related to prior experience negatively affecting attitudes toward the present experience [41]. Another possible explanation is that students' perceptions of the current smart glass–based training program might be influenced by their perception of the program itself. Since this study was intended to promote self-practice for nursing skills, a passive attitude toward training was reflected. Regardless of gender, age, and previous device experience, students' willingness and active attitude result in greater educational benefits. Further investigation regarding influencing factors on the user's perception toward smart glasses and their applications for education would offer a more comprehensive understanding for future developers.

Limitations

This study was not without limitations. First, although this was a pilot study focusing on usability and feasibility, the small sample size restricted the interpretation of some of the results. In addition, it is not possible to fully elucidate the effectiveness of the smart glass–based training program. Examining usability and feasibility, we did not thoroughly compare the effectiveness of the smart glass–based training program to other existing training programs that are prevalent in nursing education (eg, smartphone video recording of self-practice). Given the finding from this study that smart glasses can be a useful education strategy, more thoroughly demonstrating the effectiveness with future research would encourage faculties to actively incorporate such devices into their education plans. Lastly, it is questionable whether the Vuzix Blade is the best device for nursing skills training, as new smart glasses are continuously released in this

growing market. Thus, future research using a variety of smart glasses with differing specifications that reflect factors that caused discomfort and inconvenience in this study would offer valuable information for educators considering the use of smart glasses. Employing and comparing various AR presentation types (eg, 3D content, data visualization, virtual characters) and AR augmentation techniques (eg, multimodal, physical feedback, sound augmentation) are worthy of further investigation to elucidate optimal smart glasses–based practices.

Conclusion

The findings of this study suggest the use of smart glasses was a useful educational strategy for assisting self-practice of skills in nursing education. Given the benefits of timely information and hands-free operation (hands free from holding a device), participants reported positive experiences in general, including a high level of interest and appreciation for the convenience of this training program. Participants who had favorable views of this technology-enhanced education were more likely to report greater learning satisfaction, which shows great potential in transforming a previously passive attitude to an active one. Future revision reflecting the feedback from this study would effectively foster a high level of skill competency among undergraduate nursing students, engaging students in active learning and reducing the burden on faculty members.

Acknowledgments

This research was supported by a grant (20012234) of Regional Customized Disaster-Safety R&D Program funded by the Ministry of the Interior and Safety (Korea). This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (No. NRF-2019R1G1A1006737 and No. NRF-2018R1D1A1B07048247).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Snapshots of the image guide for intradermal injection, snapshots of the image guide for blood transfusion, Pearson correlation analysis among study variables, and difference of usability score (ease of use and usefulness) by gender and previous experience of augmented reality.
[DOCX File , 2121 KB - jmir_v23i3e24313_app1.docx]

References

1. The Ministry of Health and Welfare. The fourth national health plan (2016-2020). Sejong: The Ministry of Health and Welfare; 2015. URL: https://www.khealth.or.kr/board;jsessionid=QMdkJMYcxNH2rIsLt8smQIzNmdLv10lFbyXweqYyksw75vorJWGIghUJq13LJXpj.khealth-was_servlet_khealth2018?menuId=MENU00729&siteId=null [accessed 2020-06-10]
2. Aiken LH, Sloane DM, Bruyneel L, Van den Heede K, Griffiths P, Busse R, RN4CAST consortium. Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study. Lancet 2014 May 24;383(9931):1824-1830 [FREE Full text] [doi: 10.1016/S0140-6736(13)62631-8] [Medline: 24581683]
3. Hartigan I, Murphy S, Flynn AV, Walshe N. Acute nursing episodes which challenge graduate's competence: perceptions of registered nurses. Nurse Educ Pract 2010 Sep;10(5):291-297. [doi: 10.1016/j.nepr.2010.01.005] [Medline: 20167537]
4. Lin Z. Comparison of technology-based cooperative learning with technology-based individual learning in enhancing fundamental nursing proficiency. Nurse Educ Today 2013 May;33(5):546-551. [doi: 10.1016/j.nedt.2011.12.006] [Medline: 22260882]
5. Korean Accreditation Board of Nursing Education-Evaluation core basic nursing skill items. Seoul: Korean Accreditation Board of Nursing Education (KABONE); 2017. URL: http://www.kabone.or.kr/ [accessed 2020-06-20]
6. Shin SJ, Kim YK, Suh S, Jung DY, Kim Y, Yim MK. Perception survey on the introduction of clinical performance examination as part of the national nursing licensing examination in Korea. J Educ Eval Health Prof 2017;14:26 [FREE Full text] [doi: 10.3352/jeehp.2017.14.26] [Medline: 29129904]
7. Chantal L, Anik D. Preparation Guide for the Professional Examination of the Ordre des infirmières et infirmiers du Québec, 5th edition. Montreal: Ordre des infirmieres et infirmiers du Quebec; 2018.
8. Seungeun Chung, Sujin Shin. Learning Experiences of Essentials of Fundamental Nursing Skills among Nursing Students: Focused on Lab Self-training. J Qual Res 2017 May;18(1):105-113. [doi: 10.22284/qr.2017.18.1.105]
9. Hyo-ju J. A Content Analysis of the Nursing College Student's Experience in the Lab Self-training. Asia-Pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology 2019 Dec;9(12):757-766. [doi: 10.35873/ajmahs.2019.9.12.067]
10. Paik HJ. Educational evaluation of competency in nursing skills through open laboratory self-directed practice. J of Korean Acad Fundam Nurs 2004;11(1):13-20.

11. Kim SY. The Effects of Imagery Training on Outcomes of Clinical Competence, Voluntary Practice Time and Nursing Skills. The Journal of Korean Academic Society of Nursing Education 2017 Feb 28;23(1):76-85. [doi: 10.5977/jkasne.2017.23.1.76]
12. Larkin J, Simon H. Why a Diagram is (Sometimes) Worth Ten Thousand Words. Cognitive Science 1987 Jan;11(1):65-100 [FREE Full text] [doi: 10.1016/S0364-0213(87)80026-5]
13. Delany C, Golding C. Teaching clinical reasoning by making thinking visible: an action research project with allied health clinical educators. BMC Med Educ 2014 Jan 30;14:20 [FREE Full text] [doi: 10.1186/1472-6920-14-20] [Medline: 24479414]
14. Naps TL, Rößling G, Almstrum V, Dann W, Fleischer R, Hundhausen C, et al. Exploring the role of visualization and engagement in computer science education. SIGCSE Bull 2003;35(2):131-152. [doi: 10.1145/782941.782998]
15. Zweifach S, Triola M. Extended Reality in Medical Education: Driving Adoption through Provider-Centered Design. Digit Biomark 2019;3(1):14-21 [FREE Full text] [doi: 10.1159/000498923] [Medline: 32095765]
16. Aebersold M, Voepel-Lewis T, Cherara L, Weber M, Khouri C, Levine R, et al. Interactive Anatomy-Augmented Virtual Simulation Training. Clin Simul Nurs 2018 Feb;15:34-41 [FREE Full text] [doi: 10.1016/j.ecns.2017.09.008] [Medline: 29861797]
17. Huang CY, Thomas JB, Alismail A, Cohen A, Almutairi W, Daher NS, et al. The use of augmented reality glasses in central line simulation: "see one, simulate many, do one competently, and teach everyone". Adv Med Educ Pract 2018;9:357-363 [FREE Full text] [doi: 10.2147/AMEP.S160704] [Medline: 29785148]
18. Rochlen L, Levine R, Tait A. First-Person Point-of-View-Augmented Reality for Central Line Insertion Training: A Usability and Feasibility Study. Simul Healthc 2017 Feb;12(1):57-62 [FREE Full text] [doi: 10.1097/SIH.0000000000000185] [Medline: 27930431]
19. Hanna M, Ahmed I, Nine J, Prajapati S, Pantanowitz L. Augmented Reality Technology Using Microsoft HoloLens in Anatomic Pathology. Arch Pathol Lab Med 2018 May;142(5):638-644 [FREE Full text] [doi: 10.5858/arpa.2017-0189-OA] [Medline: 29384690]
20. Kim HS. A study on relationship between stress of clinical practice and clinical competency in nursing students. Journal of Korean Public Health Nursing 2002;16(1):64-76.
21. Barsom EZ, Graafland M, Schijven MP. Systematic review on the effectiveness of augmented reality applications in medical training. Surg Endosc 2016 Oct;30(10):4174-4183 [FREE Full text] [doi: 10.1007/s00464-016-4800-6] [Medline: 26905573]
22. Chang S, Hwang G. Impacts of an augmented reality-based flipped learning guiding approach on students' scientific project performance and perceptions. Computers & Education 2018 Oct;125:226-239. [doi: 10.1016/j.compedu.2018.06.007]
23. VanMeerten N, Varma K. Exploring student engagement in an augmented reality learning game. Int J Gaming Comput Mediat Simul 2017;9(4):44-61. [doi: 10.4018/IJGCMS.2017100103]
24. Herpich F, Guarese R, Cassola A, Tarouco L. Mobile augmented reality impact in student engagement: an analysis of the focused attention dimension. 2018 Presented at: International Conference on Computational Science and Computational Intelligence (CSCI); 12-14 Dec. 2018; Las Vegas, NV, USA p. 562-567. [doi: 10.1109/CSCI46756.2018.00114]
25. Kopetz J, Wessel D, Jochems N. User-centered development of smart glasses support for skills training in nursing education. i-com 2020 2020;18(3):287-299. [doi: 10.1515/icom-2018-0043]
26. Wüller H, Behrens J, Garthaus M, Marquard S, Remmers H. A scoping review of augmented reality in nursing. BMC Nurs 2019;18:19 [FREE Full text] [doi: 10.1186/s12912-019-0342-2] [Medline: 31123428]
27. Vuzix. Vuzix Blade: See-through AR Smart Glasses powered by industry-leading waveguide optics. URL: https://www.vuzix.com/products/blade-smart-glasses [accessed 2021-01-20]
28. Kim S, Nussbaum MA, Gabbard JL. Influences of augmented reality head-worn display type and user interface design on performance and usability in simulated warehouse order picking. Appl Ergon 2019 Jan;74:186-193. [doi: 10.1016/j.apergo.2018.08.026] [Medline: 30487099]
29. Ji Y, Chung H. Effects of case-based learning on task achievement and learning satisfaction in the university class. Journal of Learner-Centered Curriculum and Instruction 2014;14(9):243-265 [FREE Full text]
30. Han A, Cho DS, Won J. A Study on Learning Experiences and Self-Confidence of Core Nursing Skills in Nursing Practicum among Final Year Nursing Students. Journal of Korean Academy of Fundamentals of Nursing 2014 May 31;21(2):162-173. [doi: 10.7739/jkafn.2014.21.2.162]
31. Colley SL. Implementing a change to a learner-centered philosophy in a school of nursing: faculty perceptions. Nurs Educ Perspect 2012;33(4):229-233. [doi: 10.5480/1536-5026-33.4.229] [Medline: 22916625]
32. Kim EJ. Nursing Students' Error and Recovery in Transfusion Simulation for Safety Competency. J Korean Acad Fundam Nurs 2015 May 30;22(2):180-189. [doi: 10.7739/jkafn.2015.22.2.180]
33. Klinker K, Wiesche M, Krcmar H. Digital Transformation in Health Care: Augmented Reality for Hands-Free Service Innovation. Inf Syst Front 2019 Jun 25;22(6):1419-1431. [doi: 10.1007/s10796-019-09937-7]
34. Glass. Glass Enterprise Edition 2: Specs. URL: https://www.google.com/glass/tech-specs/ [accessed 2021-01-20]
35. Kim JH, Park H. Effects of Smartphone-Based Mobile Learning in Nursing Education: A Systematic Review and Meta-analysis. Asian Nurs Res (Korean Soc Nurs Sci) 2019 Feb;13(1):20-29 [FREE Full text] [doi: 10.1016/j.anr.2019.01.005] [Medline: 30659927]

36. Ferguson C, Davidson PM, Scott PJ, Jackson D, Hickman LD. Augmented reality, virtual reality and gaming: an integral part of nursing. Contemp Nurse 2015;51(1):1-4. [doi: 10.1080/10376178.2015.1130360] [Medline: 26678947]
37. Wood W, Li S. The empirical analysis of technology camel. Issues in Information Systems 2005;6(2):154-160.
38. Kongaut C, Bohlin E. Investigating mobile broadband adoption and usage: A case of smartphones in Sweden. Telematics and Informatics 2016 Aug;33(3):742-752. [doi: 10.1016/j.tele.2015.12.002]
39. Li S, Glass R, Records H. The Influence of Gender on New Technology Adoption and Use–Mobile Commerce. Journal of Internet Commerce 2008 Jul 24;7(2):270-289. [doi: 10.1080/15332860802067748]
40. Dirin A, Alamäki A, Suomala J. Gender Differences in Perceptions of Conventional Video, Virtual Reality and Augmented Reality. Int J Interact Mob Technol 2019 Jun 20;13(06):93. [doi: 10.3991/ijim.v13i06.10487]
41. Hopp T, Gangadharbatla H. Novelty Effects in Augmented Reality Advertising Environments: The Influence of Exposure Time and Self-Efficacy. Journal of Current Issues & Research in Advertising 2016 May 25;37(2):113-130. [doi: 10.1080/10641734.2016.1171179]

Abbreviations

API: application programming interface
AR: augmented reality
KABONE: Korean Accreditation Board of Nursing Education
NRF: National Research Foundation of Korea
XR: extended reality

Edited by R Kukafka; submitted 14.09.20; peer-reviewed by T Piumsomboon, A Choi; comments to author 01.10.20; revised version received 13.10.20; accepted 20.01.21; published 02.03.21.

Please cite as:
Kim SK, Lee Y, Yoon H, Choi J
Adaptation of Extended Reality Smart Glasses for Core Nursing Skill Training Among Undergraduate Nursing Students: Usability and Feasibility Study
J Med Internet Res 2021;23(3):e24313
URL: https://www.jmir.org/2021/3/e24313
doi:10.2196/24313
PMID:33650975

Original Paper

Assessment of Diagnostic Competences With Standardized Patients Versus Virtual Patients: Experimental Study in the Context of History Taking

Maximilian C Fink[1], MSc; Victoria Reitmeier[1], Dr med; Matthias Stadler[2,3], Dr phil; Matthias Siebeck[1,3], Prof Dr, MME (D); Frank Fischer[2,3], Prof Dr; Martin R Fischer[1], Prof Dr, MME (Bern)

[1]Institute for Medical Education, University Hospital, LMU Munich, Munich, Germany

[2]Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany

[3]Munich Center of the Learning Sciences, Ludwig-Maximilians-Universität München, Munich, Germany

Corresponding Author:
Maximilian C Fink, MSc
Institute for Medical Education
University Hospital, LMU Munich
Pettenkoferstraße 8a
Munich, 80336
Germany
Phone: 49 089 4400 57428
Email: maximilian.fink@yahoo.com

Abstract

Background: Standardized patients (SPs) have been one of the popular assessment methods in clinical teaching for decades, although they are resource intensive. Nowadays, simulated virtual patients (VPs) are increasingly used because they are permanently available and fully scalable to a large audience. However, empirical studies comparing the differential effects of these assessment methods are lacking. Similarly, the relationships between key variables associated with diagnostic competences (ie, diagnostic accuracy and evidence generation) in these assessment methods still require further research.

Objective: The aim of this study is to compare perceived authenticity, cognitive load, and diagnostic competences in performance-based assessment using SPs and VPs. This study also aims to examine the relationships of perceived authenticity, cognitive load, and quality of evidence generation with diagnostic accuracy.

Methods: We conducted an experimental study with 86 medical students (mean 26.03 years, SD 4.71) focusing on history taking in dyspnea cases. Participants solved three cases with SPs and three cases with VPs in this repeated measures study. After each case, students provided a diagnosis and rated perceived authenticity and cognitive load. The provided diagnosis was scored in terms of diagnostic accuracy; the questions asked by the medical students were rated with respect to their quality of evidence generation. In addition to regular null hypothesis testing, this study used equivalence testing to investigate the absence of meaningful effects.

Results: Perceived authenticity (1-tailed t_{81}=11.12; P<.001) was higher for SPs than for VPs. The correlation between diagnostic accuracy and perceived authenticity was very small (r=0.05) and neither equivalent (P=.09) nor statistically significant (P=.32). Cognitive load was equivalent in both assessment methods (t_{82}=2.81; P=.003). Intrinsic cognitive load (1-tailed r=−0.30; P=.003) and extraneous load (1-tailed r=−0.29; P=.003) correlated negatively with the combined score for diagnostic accuracy. The quality of evidence generation was positively related to diagnostic accuracy for VPs (1-tailed r=0.38; P<.001); this finding did not hold for SPs (1-tailed r=0.05; P=.32). Comparing both assessment methods with each other, diagnostic accuracy was higher for SPs than for VPs (2-tailed t_{85}=2.49; P=.01).

Conclusions: The results on perceived authenticity demonstrate that learners experience SPs as more authentic than VPs. As higher amounts of intrinsic and extraneous cognitive loads are detrimental to performance, both types of cognitive load must be monitored and manipulated systematically in the assessment. Diagnostic accuracy was higher for SPs than for VPs, which could potentially negatively affect students' grades with VPs. We identify and discuss possible reasons for this performance difference between both assessment methods.

(J Med Internet Res 2021;23(3):e21196) doi:10.2196/21196

KEYWORDS

clinical reasoning; medical education; performance-based assessment; simulation; standardized patient; virtual patient

Introduction

Performance-Based Assessment With Standardized Patients and Virtual Patients

Since the turn of the millennium, performance-based assessment has become a mandatory part of medical licensure examinations in various countries [1], complementing traditional assessment formats, such as text vignettes, with methods including standardized patients (SPs) and simulated virtual patients (VPs). SPs have been used for performance-based assessment in health care since the 1960s [2]. However, VPs have only recently become more widely employed in this domain [3].

The term SPs refers to (trained) actors or real former patients who act as if they display symptoms of a disease [4]. Usually, students encounter several SPs in assessment settings to reliably measure clinical variety [5]. Performance is then scored by a trained faculty member or the SPs themselves using a rating scheme. Although we will elaborate on the specific features used for this assessment method later, it should be noted here that organizing an assessment with SPs is relatively resource intensive [6].

VPs are a type of computer simulation and typically include an authentic model of a real-world situation that can be manipulated by the participant [7]. VPs can use avatars or realistic videos with SPs as stimuli and offer varying degrees of interaction [8]. Moreover, assessment through VPs can take place automatically, and a recent study showed that such an automatic assessment corresponds well to ratings from clinician-educators [9]. The production of authentic VPs can frequently produce considerable costs above $10,000 [10]. Although the initial production of VPs is often more resource intensive than organizing SPs, this assessment method is then permanently available and fully scalable to a large audience.

Next, we summarize a conceptual framework. This framework provides, on the one hand, a precise operationalization of diagnostic competences. On the other hand, the framework includes a research agenda that summarizes essential moderators of performance that should be examined systematically in research on simulation-based assessment.

A Framework for the Assessment of Diagnostic Competences With Simulations

The framework developed by Heitzmann et al [10] to facilitate diagnostic competences with simulations operationalizes diagnostic competences in assessment settings as a disposition. This disposition encompasses the components of diagnostic knowledge, diagnostic quality, and diagnostic activities. Diagnostic knowledge includes conceptual and strategic knowledge [11]. Conceptual knowledge encompasses concepts and their relationships. Strategic knowledge comprises possible avenues and heuristics in diagnosing. Diagnostic quality consists of components' diagnostic accuracy and efficiency that can serve as major outcome measures in empirical studies. Diagnostic activities entail the actions of persons assessed during the diagnostic process, such as evidence generation by asking questions in history taking. The framework proposes that context is an important moderator in assessment. Therefore, more research on the effects of the assessment methods SPs and VPs seems to be warranted. A meta-analysis on simulation-based learning of complex skills [12] added to this framework that authenticity should also be explored as an important moderator in assessment and learning. Similarly, a meta-analysis on instructional design features in simulation-based learning indicated that certain types of cognitive load could be detrimental to performance [13]. Therefore, it could be fruitful to explore the relationship between cognitive load and diagnostic competences within SP and VP assessments.

Perceived Authenticity and Diagnostic Competences With SPs and VPs

There is a multitude of conceptualizations of authenticity. In our study, we focus on *perceived authenticity* [14] because this concept can be assessed entirely internally by learners' judgment. Other related concepts such as *thick authenticity* [15] and *fidelity* [16] can, at least to some extent, also be determined externally.

According to a factor analysis by Schubert et al [14], perceived authenticity—sometimes also called presence—comprises the facets of realness, involvement, and spatial presence. Realness describes the degree to which a person believes that a situation and its characteristics resemble a real-life context [14]. Involvement is defined as a feeling of cognitive immersion and judgment that a situation has personal relevancy [17]. Spatial presence denotes the feeling of physical immersion in a situation [14]. SPs are considered highly authentic because they are carefully trained to realistically portray symptoms and allow for natural interactions [18]. Empirical studies support this claim, reporting high values of perceived authenticity for SPs [19,20]. VPs also received rather high perceived authenticity scores in empirical studies [21] but lacked some of the features that may make SPs particularly authentic, such as high interactivity in oral conversations. Thus, VPs could potentially evoke lower perceived authenticity than SPs. Findings on the effect of authenticity on diagnostic competences are mixed. On the one hand, it has been argued that higher authenticity is associated with higher engagement and better performance [22]. On the other hand, literature reviews [23,24] that compared the relationship between perceived authenticity and clinical performance in simulation-based learning only reported minimal effects of authenticity. In addition, an empirical study [25] showed that above a certain threshold, further increases in perceived authenticity do not improve diagnostic accuracy.

Cognitive Load and Diagnostic Competences With SPs and VPs

Cognitive load theory posits that performance can be inhibited through high situational demands that stress working memory and attention [26]. The cognitive load consists of the following 3 different facets [27]: *Intrinsic* load results from the interplay between certain topics and materials and the assessed person's expertise. *Extraneous load* is created exclusively by characteristics of the assessment environment that strain memory and attention without being necessary for performance. *Germane load* refers to the cognitive load created through the assessed person's cognitive processes, including schema construction and abstraction. Intrinsic and extraneous cognitive loads are considered additive and can inhibit performance in complex tasks [27]. Germane load, however, is theorized to bolster performance [27]. A few primary studies from medical education have already contrasted the cognitive load of different assessment methods and reported their relationship with diagnostic competences. Dankbaar et al [28] demonstrated that intrinsic and germane cognitive loads were higher for a group learning emergency skills with a simulation game than for a group learning with a text-based simulation. Extraneous load did not differ between these groups, and none of the groups differed in performance. Haji et al [29] compared surgical skills training with less complex and more complex simulation tasks. The total cognitive load was higher in the more complex simulation than in the less complex simulation, and cognitive load was negatively associated with performance. As a result of these findings, we can conclude that SPs and VPs generally do not differ in different facets of cognitive load if the assessment methods are of equal complexity, and the main characteristics related to the facets are similar. The literature summarized earlier also shows that intrinsic and extraneous cognitive loads are negatively associated with diagnostic competences.

Assessment Method and Diagnostic Competences

Before we discuss diagnostic accuracy and evidence generation—2 important aspects of diagnostic competences—it should be noted that diagnostic competences are only a part of the broader concept of clinical reasoning. Clinical reasoning emphasizes the process of diagnosing and encompasses the full process of making clinical decisions, including the selection, planning, and reevaluation of a selected intervention [30]. In line with the conceptual framework by Heitzmann et al [10] for facilitating diagnostic competences, *diagnostic accuracy* denotes the correspondence between the learner's diagnoses and the solutions determined by experts for the same cases. According to this framework, *evidence generation* (ie, actions related to the gathering of data in a goal-oriented way) is also an important quality criterion for the diagnostic process and a crucial aspect of diagnostic competences.

Diagnostic Accuracy

Currently, there are only a few studies in the health care domain that contrast assessments using VPs and SPs directly in one experiment. Edelstein et al [1] investigated assessments with SPs and computer-based case simulations in advanced medical students using a repeated measures design. A moderate positive correlation was found between diagnostic accuracy in the two assessment formats that used different cases. Guagnano et al [31] examined SPs and computer-based case simulations in a medical licensing exam. Participants first completed the computer-based case simulations and then completed the SPs. The two assessment methods correlated positively with each other. Hawkins et al [32] compared the assessment of patient management skills and clinical skills with SPs and computer-based case simulations in a randomized controlled trial. Participating physicians completed both assessment methods, and a positive correlation of diagnostic accuracy with both assessment methods was reported. Outside the health care domain, a meta-analysis of studies from different domains reported a robust modality effect for students in problem-solving tasks. Students who solved problems presented in the form of illustrations accompanied by text were more successful than students who solved problems presented merely in text form [33]. Similarly, it seems reasonable to assume that one assessment method could lead to higher diagnostic accuracy than the other assessment method because of its different characteristics. The described findings from the health care domain tentatively indicate that SPs and VPs could result in relatively equivalent diagnostic accuracy. Such a finding would contradict the modality effect reported in other domains.

Evidence Generation

Comparable empirical studies on evidence generation for SPs and VPs are lacking. Nevertheless, we can assume that the quantity of evidence generation should be higher for SPs than for VPs. The main reason for this is that students can ask questions of SPs more quickly orally than by selecting questions from a menu of options with VPs. Apart from this difference in evidence generation between the 2 assessment methods, the relationships between evidence generation and diagnostic accuracy are interesting. The relationship between the quantity of evidence generation and diagnostic accuracy is relatively complex. The ideal amount of evidence generation may depend strongly on the case difficulty, the diagnostic cues contained in the evidence, and learner characteristics. For these reasons, the framework by Heitzmann et al [10] for facilitating diagnostic competences argues that the sheer quantity of evidence generation is not a dependable quality criterion for the diagnostic process. However, the quality of evidence generation is hypothesized by Heitzmann et al [10] to be a rather dependable quality criterion for the diagnostic process. This agrees with the literature, as we know from studies on SPs using observational checklists that the quality of evidence generation is positively associated with diagnostic accuracy [34]. Moreover, one study with specialists in internal medicine and real patients demonstrated that asking specific questions in history taking correlated positively with clinical problem solving [35].

Study Aim, Research Questions, and Hypotheses

We aim to compare the perceived authenticity, cognitive load, and diagnostic competences in SPs and VPs. We also aim to examine the relationships of perceived authenticity, cognitive load, and quality of evidence generation with diagnostic accuracy. Thus, we address the following 3 research questions: To what extent does perceived authenticity differ across the 2

assessment methods, and how is it associated with diagnostic accuracy (RQ1)? We hypothesize that SPs induce higher perceived authenticity than VPs (H1.1). Moreover, we expect to be able to demonstrate with equivalence tests for correlations (given in the *Statistical Analyses* section) that perceived authenticity is not associated meaningfully with diagnostic accuracy (H1.2). Next, is cognitive load equivalent for SPs and VPs, and how is it related to diagnostic accuracy (RQ2)? We assume to find equivalent cognitive load for SPs and VPs (H2.1). Moreover, we expect that intrinsic and extraneous loads are negatively related to diagnostic accuracy (H2.2-H2.3). To what extent are the diagnostic competences components diagnostic accuracy, quantity of evidence generation, and quality of evidence generation equivalent or differ for SPs and VPs, and how are they related to each other (RQ3)? We hypothesize that SPs and VPs evoke equivalent diagnostic accuracy (H3.1). In addition, we assume that the quantity of evidence generation is higher for SPs than for VPs (H3.2). We also expect that the quality of evidence generation is positively related to diagnostic accuracy (H3.3).

Methods

Participant Characteristics and Sampling Procedures

A sample of 86 German medical students (with a mean age of 26.03 years, SD 4.71) made up the final data set. This sample consisted of 63% (54/86) females and 37% (32/86) males. Medical students in years 3-6 of a 6-year program with a good command of German were eligible. Medical students in years 3-5 (44/86, 51%) were considered novices, as they were still completing the clinical part of the medical school. Medical students in year 6 (42/86, 49%) were regarded as intermediates as they had passed their second national examination and worked full time as interns in a medical clinic or practice. We provide a detailed overview of participant characteristics across all conditions and a CONSORT (Consolidated Standards of Reporting Trials)–style diagram of participant flow in Multimedia Appendix 1.

We collected data from October 20, 2018, to February 20, 2019, in the medical simulation center of the University Hospital, LMU Munich. We recruited participants via on-campus and web-based advertising. Participants were randomly assigned to conditions by the first author by drawing a pin code to log in to an electronic learning environment without knowing the condition assigned to the pin. In the final data collection sessions, the conditions were filled by the first author with random participants from specific expertise groups (novices vs intermediates). This procedure was applied to achieve a comparable level of expertise in all conditions. As expected, the proportion of participants from different expertise groups did not differ across conditions (χ^2_3=0.2; *P*=.99).

Research Design

The study used a repeated measures design with assessment method (SPs vs VPs) as the key factor. In addition, we varied the between-subjects factor case group (CG) order and assessment method order. In total, students encountered 6 different cases. We provide an overview of the experiment in Table 1. Details of the succession through cases and medical content in the experimental conditions are provided in Table 2. We attempted to ensure similar topics and difficulty for both CGs by conducting an expert workshop and adapting cases based on the experts' feedback as part of creating the experimental materials.

Table 1. General overview of the experiment.

Part of the experiment	Activity or test	Duration (min)
Pretest	Briefing	10
	Conceptual knowledge test	40
	Strategic knowledge test	40
Break	—[a]	10
Assessment phase I (cases 1-3)	VPs[b] or SPs[c]	70
Break and change of modality	—	5
Assessment phase II (cases 4-6)	VPs or SPs	70
Posttest and debriefing	Working memory test	15
	End-debriefing	5

[a]No activity or test takes place.

[b]VP: virtual patient.

[c]SP: standardized patient.

Table 2. Succession through cases and medical content in the experimental conditions[a,b].

Cases	Condition 1A	Condition 1B	Condition 2A	Condition 2B
1-3	CG[c] A (SPs[d])	CG B (VPs[e])	CG B (SPs)	CG A (VPs)
4-6	CG B (VPs)	CG A (SPs)	CG A (VPs)	CG B (SPs)

[a]Case group A: (1) pulmonary embolism with lymphoma, (2) congestive heart failure with atrial fibrillation, and (3) hyperventilation tetany caused by a panic attack.

[b]Case group B: (1) pulmonary embolism with coagulation disorder, (2) community-acquired pneumonia, and (3) hypertrophic obstructive cardiomyopathy.

[c]CG: case group.

[d]SP: standardized patient.

[e]VP: virtual patient.

Procedure and Materials

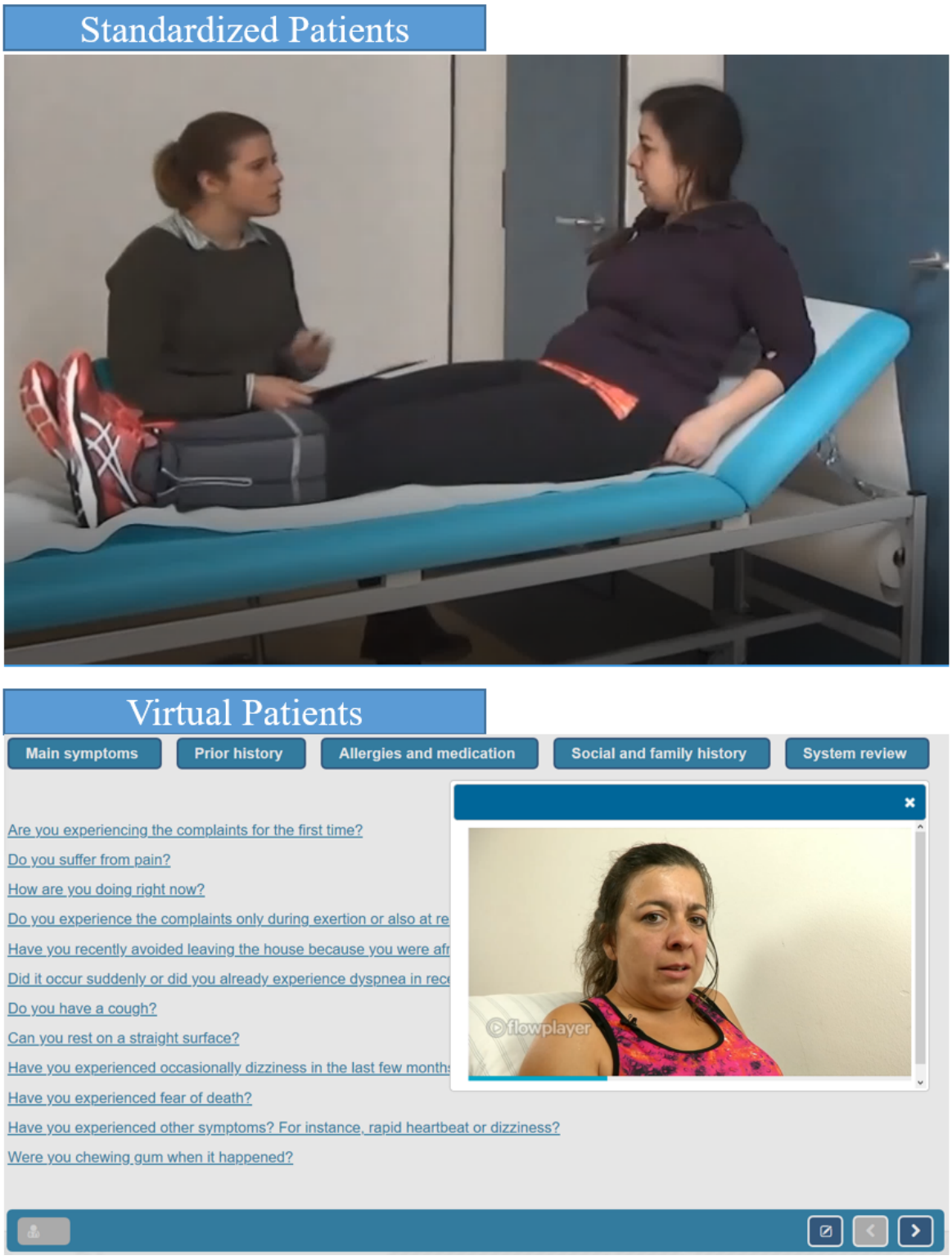
Participants completed a pretest of conceptual knowledge and strategic knowledge at the beginning of the experiment. Afterward, participants took part in the assessment phase, solving the first 3 cases with SPs and the next 3 cases with VPs or vice versa. All cases were drafted by a specialist in general practice and evaluated positively by an expert panel. The cases were not adapted from real clinical cases but based on cases from textbooks and symptoms reported in guidelines. A short familiarization phase preceded each assessment phase and included a motivational scale. For all cases in both assessment methods, assessment time was held constant at 8 minutes and 30 seconds for history taking and 5 minutes for writing up a diagnosis for the case in an electronic patient file. At the end of the experiment, participants were debriefed. A more detailed overview of the procedure can be found in Multimedia Appendix 2.

Assessment with SPs was conducted in a simulated emergency room. All SPs were (semi-) professional actors who were financially compensated; most had previous experience working in an SP program. All SPs were extensively trained by an acting coach and a physician, memorized their symptoms and scripts, and were not aware of their patient's diagnosis. Participants first received prior information (eg, electrocardiogram and lab results) and presentation of the chief complaint for each case. Next, participants formulated and asked questions independently, and the SPs responded. The interaction was recorded on a video. After each case, the participants completed a patient file, including measures of diagnostic accuracy and other scales. A screenshot of this assessment method is provided in Figure 1.

The assessment with the VPs was carried out in a simulated assessment environment in a computer room. First, participants received prior information and a video with a chief complaint for each case. The participants then selected questions independently from a menu with up to 69 history-taking questions. The VP's answer was streamed as a video, including a recorded response by an actor. After each case, the participants completed a patient file, including a measure of diagnostic accuracy and other scales. A screenshot of this assessment method is provided in Figure 1.

The VPs, patient file, and other measures were implemented in the electronic assessment environment CASUS [36]. The questions provided for the VPs were based on a structural and topical analysis of history-taking forms by Bornemann [37] and are displayed in Multimedia Appendix 3. According to this analysis, physician questions in history taking can fall under the 5 categories of main symptoms, prior history, allergies and medication, social and family history, and system review. Participants with SPs received empty history-taking forms for all cases and time to formulate possible history-taking questions during the familiarization phase, at which point participants in the VPs only read all questions from the menu. Without this additional structuring support in the SP condition, the participants in the VP condition would have received additional support in the form of a list of questions in the menu.

Figure 1. History-taking with standardized patients and virtual patients.



Measures and Covariates

Perceived Authenticity

Perceived authenticity was operationalized as a construct with the 3 dimensions of realness, involvement, and spatial presence [14]. All 3 authenticity scales used a 5-point scale ranging from (1) *disagree* to (5) *agree* and were taken from multiple validated questionnaires [14,38-40]. The items were slightly adapted to simulation-based assessment and are included in Multimedia Appendix 4. A combined score for all 3 dimensions was built by calculating the mean. This scale achieved a reliability of Cronbach α=.88.

Cognitive Load

The cognitive load scale by Opfermann [41] used in this study assessed the extraneous cognitive load with 3 items and germane and intrinsic cognitive loads with 1 item each. A 5-point scale from (1) *very easy*, (2) *rather easy*, (3) *neutral*, (4) *rather hard*,

to (5) *very hard* was used. The scale is included in Multimedia Appendix 4. A combined score for all 3 facets was built by calculating the mean. This scale achieved a reliability of Cronbach α=.88.

Motivation, Diagnostic Knowledge, and Other Control Variables

We assessed motivation as a control variable because it could differ between assessment methods and potentially affect performance. The expectancy component of motivation was assessed with a 4-item, 7-point scale adapted from Rheinberg et al [42]. The motivation expectancy scale ranged from (1) strongly disagree to (7) strongly agree. The value component of motivation was measured with a 4-item, 5-point scale based on a questionnaire by Wigfield [43]. The motivation value scale ranged from (1) strongly disagree to (5) strongly agree. The full scales are provided in Multimedia Appendix 4. Diagnostic knowledge was also measured in this study but later not taken into account in the analyses because it was similar in VPs and SPs because of the repeated measures design. We measured diagnostic knowledge using a conceptual and strategic knowledge test. Both types of knowledge have been identified as predictors of clinical reasoning [44]. The maximum testing time was set to 40 minutes per test. More details on both diagnostic knowledge tests are reported in Multimedia Appendix 4. Apart from this, demographic data were collected, including participants' sex, age, and expertise (year of medical school).

Diagnostic Competences

Diagnostic Accuracy

Diagnostic accuracy was assessed based on the answer to the prompt "Please choose your final diagnosis after history taking" from a long menu containing 239 alternative diagnoses. Two physicians created a coding scheme for scoring diagnostic accuracy in all cases (Multimedia Appendix 4). To do that, the physicians rated all 239 alternative diagnoses for all cases and resolved the disagreements until they reached full agreement. One of the physicians was a specialist in general practice who also drafted the cases. The other physician was a board-certified doctor familiar with medical assessment through her dissertation. The latter physician, who is also the second author of this paper, then scored diagnostic accuracy based on the coding scheme: 1 point was allocated for the designated correct answer, 0.5 point for a partially correct answer, and 0 point for an incorrect answer. Due to having only 1 rater to score the diagnostic accuracy with the comprehensive coding scheme, a reliability estimate cannot be reported. However, this is also not necessary because the exact diagnostic accuracy score for all selectable diagnoses included in the electronic assessment environment was determined upfront in the coding scheme.

Evidence Generation

The second author classified the quality of evidence generation by determining the essential questions relevant for the correct diagnosis for each VP case (the coding scheme is given in Multimedia Appendix 4). This process took part before looking at the experimental data. All solutions were discussed with a specialist in general practice, and all disagreements were resolved. Student assistants transcribed all utterances recorded in the videos of the SP encounters, and the electronic assessment environment stored all selected questions during the VP encounters. The *R* scripts automatically classified the log data from the VPs using the coding scheme. Student assistants had no medical background and were trained by the second author to code the transcripts from the SP encounters. This task mainly implied recognizing the intent of history-taking questions and linking them, if possible, to the most similar question in the coding scheme. After training the raters, 20% of this complex and extensive SP data were coded by 2 raters to check interrater agreement. This data set encompassed SP data from 18 of the 86 participants of our study with all three SP cases in which the participants took part. Fleiss κ=0.74 demonstrated that agreement was substantial, and the rest of the data were coded by the same raters individually. The score for quantity of evidence generation corresponded to the total number of questions posed for each case. To calculate the score for quality of evidence generation for each case, we counted the number of relevant questions posed and divided this score by the number of relevant questions that could potentially be posed.

Scale Construction

Diagnostic accuracy and evidence generation scales for each assessment method and combining the 2 methods were built by calculating the mean of the included cases. Case 1 in CS A was excluded from all analyses because of high difficulty (mean diagnostic accuracy 0.05, SD 0.18).

Statistical Analyses

This study answers the proposed research questions using traditional null hypothesis significance testing (NHST) and equivalence testing. In contrast to NHST, equivalence testing can be used to investigate "whether an observed effect is surprisingly small, assuming that a meaningful effect exists in the population" [45]. For this type of test, first, the smallest effect size of interest, that is, the threshold for a meaningful effect, is specified based on the literature. The null hypothesis that the effect is more extreme than the smallest effect size of interest is then investigated. To do this, 2 separate 1-sided tests (TOST; eg, *t* tests) are conducted [46]. These tests examine whether the observed effect is more extreme than the specified smallest effect size of interest. If both 1-sided tests are significant, the null hypothesis that there is a meaningful effect that is more extreme than the smallest effect size of interest is rejected. Thus, equivalence is supported. For more convenient reporting, only the *t* test with a higher P value is reported. In cases in which equivalence cannot be supported, NHST is performed for follow-up analyses.

All statistical analyses were performed using *R* version 3.6.1 [47]. The TOST procedure and the corresponding package TOSTER [45] were used to conduct the equivalence tests. In all statistical analyses, the alpha level was set to 5%; 1-tailed tests were used where applicable. The Bonferroni-Holm method [48] was used to correct P values for multiple comparisons in post hoc and explorative tests.

For all equivalence tests, the smallest effect size of interest was determined based on the discussed literature. For H1.2 and related post hoc tests, the smallest effect size of interest was set

to be more extreme than r=±0.20, which corresponds to the effect size of small but meaningful correlations typically encountered in the social sciences [49]. For H2.1 and related post hoc tests, a meaningful effect was determined as an effect of Cohen d=0.35. This effect size lies between a small effect (Cohen d=0.20) and a medium effect (Cohen d=0.50) [49] and occurs frequently in the social sciences. For H3.1, we determined that a meaningful effect exists in the case of a difference of ±0.125 points in diagnostic accuracy. This was based on supposing a pass cutoff of 0.50 for diagnostic accuracy (ranging from 0 to 1) and setting 4 equal intervals for the hypothetical passing grades A-D.

Power Analysis

We conducted a priori power analysis for dependent samples t tests (H1.1 and H3.2). This power analysis was based on a small to medium effect of Cohen d=0.30, 2-tailed testing, an error probability of 5%, and 80% power, resulting in a targeted sample of 90 participants. Moreover, we carried out a priori power analyses for 1-tailed correlations with r=±0.25, an error probability of 5%, and 80% power (H2.2-H2.3 and H3.3). This power analysis resulted in a planned sample size of 95 participants. A post hoc power analysis for the main equivalence test (H3.1) with 86 participants, the observed effect of Cohen d=0.26, and an error probability of 5% resulted in a power of 78%. All power analyses were conducted using G*Power software [50].

Results

Descriptive Statistics and Analysis of Control Variables

Descriptive statistics are provided in Table 3. The perceived authenticity variables were rated as very high for SPs and relatively high for VPs. Cognitive load variables were reported to be moderate in both assessment methods. The average diagnostic accuracy was medium. The quantity of evidence generation was higher for SPs than for VPs. The quality of evidence generation was medium for both assessment methods. Motivational variables were rated rather highly for both SPs and VPs. A post hoc comparison showed that the value aspect of motivation was higher for SPs than for VPs (2-tailed t_{83}=2.89; P=.01; Cohen d=0.31), whereas the expectancy aspect did not differ between assessment methods (2-tailed t_{83}=0.44; P=.66; Cohen d=0.05). Participants demonstrated slightly above medium performance on the conceptual and strategic knowledge tests. Multimedia Appendix 5 provides an additional visualization of the results using boxplots and bee swarm plots.

Table 3. Descriptive statistics.

Variable	Both methods, mean (SD)	SPs[a], mean (SD)	VPs[b], mean (SD)
Perceived authenticity[c]	3.62 (0.67)	4.02 (0.67)	3.23 (0.84)
Realness[c]	3.71 (0.79)	4.13 (0.74)	3.28 (1.07)
Involvement[c]	3.82 (0.66)	4.03 (0.73)	3.61 (0.83)
Spatial presence[c]	3.35 (0.80)	3.89 (0.83)	2.80 (1.05)
Cognitive load[c]	2.88 (0.61)	2.88 (0.74)	2.90 (0.69)
Intrinsic load[c]	3.18 (0.68)	3.20 (0.78)	3.14 (0.80)
Extraneous load[c]	2.84 (0.65)	2.82 (0.79)	2.87 (0.76)
Germane load[c]	2.74 (0.76)	2.73 (0.88)	2.76 (0.84)
Diagnostic competences			
Diagnostic accuracy[d]	0.46 (0.18)	0.51 (0.28)	0.41 (0.24)
Quantity of evidence generation	22.26 (4.88)	29.01 (8.03)	17.34 (4.21)
Quality of evidence generation[d]	0.40 (0.11)	0.37 (0.18)	0.43 (0.13)
Control variables			
Motivation expectancy aspect[e]	5.07 (0.91)	5.10 (0.88)	5.05 (1.08)
Motivation value aspect[c]	4.44 (0.51)	4.54 (0.54)	4.34 (0.67)
Conceptual knowledge[d]	0.65 (0.14)	—[f]	—
Strategic knowledge[d]	0.66 (0.15)	—	—

[a]SP: standardized patient.

[b]VP: virtual patient.

[c]Scale range: 1-5.

[d]Scale range: 0-1.

[e]Scale range: 1-7.

[f]Knowledge was assessed before taking part in SPs and VPs.

Perceived Authenticity and Diagnostic Accuracy (RQ1)

A paired sample *t* test demonstrated that in line with hypothesis H1.1, perceived authenticity was considered higher for SPs than VPs in terms of the combined score (1-tailed t_{81}=11.12; *P*<.001; Cohen *d*=1.23). Post hoc tests showed that this was also the case for realness (t_{80}=8.83; *P*<.001; Cohen *d*=0.98), involvement (t_{81}=4.60; *P*<.001; Cohen *d*=0.51), and spatial presence (t_{79}=10.65; *P*<.001; Cohen *d*=1.19). Our expectation in H1.2 was that perceived authenticity would not be meaningfully associated with diagnostic accuracy. The TOST procedure for correlations showed that the relationship between diagnostic accuracy and the combined perceived authenticity score (*r*=0.05; *P*=.09) was outside the equivalence bounds of a meaningful effect of *r*=±0.20. Post hoc equivalence tests demonstrated that this also holds for the relationship of diagnostic accuracy with realness (*r*=0.03; *P*=.06), involvement (*r*=0.07; *P*=.11), and spatial presence (*r*=0.05; *P*=.08). Reanalyzing these correlations with regular 1-tailed NHST tests also yielded nonsignificant results for the combined score (*P*=.32), realness (*P*=.39), involvement (*P*=.28), and spatial presence (*P*=.33). These results mean that there is neither evidence for the absence of meaningful correlations nor evidence for significant correlations. These inconclusive findings may stem from the lack of statistical power because of the relatively small sample size [45].

Cognitive Load and Diagnostic Accuracy (RQ2)

We hypothesized in H2.1 that we would find equivalent cognitive load scores for SPs and VPs. Equivalence testing with the TOST procedure for paired samples indicated that for both assessment methods, the scores for combined cognitive load (t_{82}=2.81; *P*=.003) were significantly within the equivalence bounds of an effect of Cohen *d*=0.35. Adjusted post hoc equivalence tests showed that this is also the case for intrinsic load (t_{82}=−2.47; *P*=.008), extraneous load (t_{82}=2.55; *P*=.01), and germane load (t_{82}=2.64; *P*=.01). We expected in H2.2-H2.3 to uncover negative correlations between diagnostic accuracy and intrinsic cognitive load and extraneous load. As assumed, intrinsic cognitive load (1-tailed *r*=−0.30; *P*=.003) and extraneous load (1-tailed *r*=−0.29; *P*=.003) correlated negatively with the combined score for diagnostic accuracy. Adjusted explorative follow-up analyses showed that germane load (*r*=−0.25; *P*=.010) and the total score for cognitive load

($r=-0.31$; $P=.004$) also correlated negatively with the combined score for diagnostic accuracy.

Assessment Method and Diagnostic Competences (RQ3)

Diagnostic Accuracy

In H3.1, we hypothesized finding equivalent diagnostic accuracy scores for SPs and VPs. H3.1 was first examined by applying a paired samples TOST procedure. According to our data, we cannot reject hypothesis H3.1 that a difference in diagnostic accuracy of at least ±0.125 points (1 grade) exists between the 2 assessment methods ($t_{85}=-0.60$; $P=.28$). A follow-up 3-way mixed design analysis of variance demonstrated that neither the CG order nor the assessment method order ($F_{3,82}=2.49$; $P=.12$; $\eta^2=0.03$, respectively, $F_{3,82}=0.02$; $P=.88$; $\eta^2=0.01$) had a significant effect on diagnostic accuracy. The assessment method itself, however, had a significant main effect ($F_{3,82}=6.30$; $P=.01$; $\eta^2=0.07$), indicating that diagnostic accuracy was higher for SPs than for VPs. The finding that diagnostic accuracy was higher for SPs than for VPs also corresponds to the result of a paired sample *t* test (2-tailed $t_{85}=2.49$; $P=.01$; Cohen $d=0.27$).

Evidence Generation

H3.2 that students display an increased quantity of evidence generation with SPs than with VPs was supported (1-tailed $t_{69}=12.26$; $P<.001$; Cohen $d=1.47$). However, in an explorative follow-up analysis, we found no evidence that the *quantity* of evidence generation was related to diagnostic accuracy (1-tailed $r=0.11$; $P=.15$). This finding holds equally for SPs ($r=-0.09$; $P=.76$) and VPs ($r=-0.10$; $P=.82$). Moreover, H3.3 that the *quality* of evidence generation is positively related to diagnostic accuracy in both assessment methods was not supported (1-tailed $r=0.18$; $P=.05$). Corrected post hoc analyses showed, however, that the quality of evidence generation was positively related to diagnostic accuracy for VPs ($r=0.38$; $P<.001$); this finding did not hold for SPs ($r=0.05$; $P=.32$). Additional post hoc exploratory analyses revealed that the quality of evidence generation was higher for VPs than for SPs (2-tailed $t_{74}=-2.47$; $P=.02$; Cohen $d=0.29$).

Discussion

Principal Findings

With regard to perceived authenticity, our results showed that SPs and VPs achieved high scores on all 3 dimensions of realness, involvement, and spatial presence. Despite this high level of perceived authenticity in both assessment methods, perceived authenticity was higher for SPs than for VPs on all 3 dimensions. This finding is in line with the literature, which has long claimed that SPs achieve a very high level of perceived authenticity [18-20]. Other studies on perceived authenticity have so far focused on comparing formats such as SPs, video presentations, and text vignettes and different levels of authenticity within VPs [21]. Our study extends this literature by directly comparing SPs and VPs with respect to 3 frequently used perceived authenticity variables. This comparison seems particularly relevant, as both assessment formats are becoming increasingly popular. Our findings on the relationship between perceived authenticity and diagnostic accuracy are mixed. The equivalence test on correlations was not significant; therefore, we could not confirm the hypothesis that perceived authenticity is not meaningfully associated with diagnostic accuracy. However, a regular correlation between perceived authenticity and diagnostic accuracy that was calculated afterward was close to 0. Taken together, these findings of nonequivalence and nonsignificance indicate that we did not have sufficient power to draw a conclusion [45]. Nevertheless, we have found some indication that the correlation between perceived authenticity and diagnostic competences is rather small. This finding is in accordance with literature reviews [23,24], which reported small correlations between perceived authenticity and performance.

With regard to cognitive load, we found that the combined score is equivalent for SPs and VPs that use the same clinical cases. This finding substantiates the literature suggesting that cognitive load depends mainly on task complexity [29]. Moreover, the fact that the extraneous load was equivalent for SPs and VPs indicates that user interaction through a software menu does not substantially increase cognitive load. This finding is important because decreasing the cognitive load by allowing for user input using natural language processing [21] is still highly expensive. Our study also adds to the literature that the level of cognitive load is similar in SPs and VPs as assessment methods if the different types of cognitive load are systematically controlled for during the design process. In addition, we demonstrated that intrinsic and extraneous cognitive loads correlate negatively with diagnostic accuracy. The finding on intrinsic cognitive load corroborates that the interplay between materials and the assessed person's expertise is associated with performance. The finding on extraneous cognitive load shows that unnecessary characteristics of the assessment environment can strain memory and attention and be detrimental to performance in assessment settings. Together, these findings fit well with the literature, which has repeatedly reported negative effects of intrinsic and extraneous cognitive loads on complex problem solving in medical education [27] and other domains [51]. Our study unveils that a negative relationship between intrinsic and extraneous cognitive loads and performance in a simulation-based measure of diagnostic competences already shows when overall cognitive load is medium on average.

Our study found no evidence that diagnostic accuracy was equivalent for SPs and VPs. In contrast, higher diagnostic accuracy was achieved for SPs than for VPs. The small number of studies comparing both assessment methods so far [1,31,32] have reported medium correlations, not taking into account different case content or testing time. Using the TOST procedure as a novel methodological approach, our study contributes to the literature by finding that grading was not equivalent, as participants received a better hypothetical grade when the simulation-based assessment was administered with SPs than with VPs. On the one hand, we cannot rule out that this finding may be explained by additional support from the actors in the SP assessment. To avoid and mitigate such an effect, actors were trained by an acting coach and a physician, memorized their symptoms and scripts, and did not know the diagnosis of

their case. Moreover, student assistants screened all SP assessments, and no additional systematic support by actors was discovered. On the other hand, this finding can be explained by the lower appraisal of motivational value and the lower quantity of evidence generation reported for VPs. Participants solving VP cases may thus have been less engaged and may have collected a smaller number of important diagnostic cues that supported their diagnostic process.

Contrary to our expectations, the quality of evidence generation was not positively correlated with the *combined* diagnostic accuracy score. Closer inspection of the data revealed that the quality of evidence generation was positively correlated with diagnostic accuracy in *VPs*. This confirmed relationship is in line with the theoretical assumptions of Heitzmann et al [10]. In *SPs*, however, the quality of evidence was not correlated with diagnostic accuracy. This finding contradicts the theoretical assumptions of Heitzmann et al [10] and empirical results from studies using observational checklists with SPs [34] and real patients [36]. There are 2 explanations for these conflicting findings. First, the quality of evidence generation was, as an exploratory follow-up *t* test indicated, higher in VPs than in SPs. This higher quality of evidence generation could have been caused by a slightly different process of history taking in both assessment methods. Participants working with VPs selected questions from a menu. In contrast, participants working with SPs formulated questions during history taking freely. Second, SPs could have offered additional support to assessed persons who displayed a low quality of evidence generation, whereas VPs reacted in a completely standardized way to all assessed persons.

Limitations

One methodological limitation of our study might be the low statistical power for the analysis of hypothesis H1.2 and related post hoc analyses that addressed the relationship between the perceived authenticity variables and diagnostic accuracy. This lack of statistical power can primarily be attributed to our investigation of whether a correlation of $r=\pm0.20$ or more extreme exists. As recommended by Lakens [46], the smallest effect size of interest was selected based on findings from the literature. Specifying the smallest effect size of interest to be larger would have increased power but not have contributed findings from a valuable equivalence test to the literature. This is the case because the literature already assumes a small effect size [23,24].

One theoretical limitation of the study is that the results on perceived authenticity may not generalize without restrictions to other related concepts of authenticity. Shaffer et al [15] argue that thick authenticity consists of four different aspects. An authentic task, situation, or material should (1) exist in real life, (2) be meaningful, (3) allow the learner to engage in professional activities of the discipline, and (4) be conducted rather similar in instruction and assessment. The authors assume that thick authenticity can only be achieved when all aspects of authenticity are adequate and that VPs could potentially achieve similar authenticity to SPs. Hamstra et al [16] proposed distinguishing fidelity using the terms physical resemblance and functional task alignment. The authors report weak evidence for the relationship between physical resemblance and performance, and strong evidence for the relationship between functional task alignment and performance. In our study, the concepts of thick authenticity and fidelity were not measured for two reasons. First, these concepts can, to some extent, only be judged externally by experts. Second, the repeated measures design of the study forced us to keep aspects such as thick authenticity, physical resemblance, and functional task alignment as similar as possible in SPs and VPs. Nevertheless, we believe that the relationship between different authenticity concepts and diagnostic competences still requires further research. Future studies should attempt to untangle the relationship between different authenticity concepts and diagnostic competences by measuring these systematically.

Conclusions

Our findings on the relationship between perceived authenticity and diagnostic accuracy contribute to the debate on the costs and benefits of perceived authenticity in performance-based assessments. These results relativize the importance of perceived authenticity in assessment. Increasing the perceived authenticity of assessment methods above a certain necessary threshold and thus raising their costs [23] does not seem to be of much benefit. Such spending could potentially squander a large share of the medical education budget [52] that could be put to more valuable use. Our results on cognitive load highlight its importance as a process variable in assessment settings. Performance-based assessment should thus attempt to reduce extraneous load and control for intrinsic load to measure performance in a standardized way that is still close to clinical practice [53].

Finally, the findings on diagnostic competences have some practical implications if VPs are used as an alternative to SPs in assessment. In particular, we found that VPs could lead to lower diagnostic accuracy scores than SPs, which could, in turn, negatively affect students' grades. There are 2 different mechanisms that could explain this finding: assessment with SPs could overestimate true performance or assessment with VPs could underestimate true performance. In accordance with SPs overestimating performance, we could not rule out additional support from the actors. In fact, the low, nonsignificant correlation between the quality of evidence generation and diagnostic accuracy in SPs, together with the higher diagnostic accuracy in SPs, could indicate that actors provided some additional support (eg, to participants who displayed low quality of evidence generation). Careful training [54] and screening thus seem to be of great importance to avoid additional support from actors during SP assessment to match the high level of standardization that VPs provide. The mechanism of possible underestimation of performance with VPs could be substantiated by the lower motivational value and quantity of evidence generation discovered for VPs. We suggest taking the following measures: students could be motivated additionally in VP assessment by more interactive environments (eg, using natural language processing) or providing automated elaborated feedback directly after the assessment. Moreover, the assessment time can be extended when menu-based VPs are used in practice. This way, the quantity of evidence generation could be raised to a level similar to that in the SP assessment.

Acknowledgments

The authors would like to thank Hannah Gerstenkorn, who developed the case vignettes. In addition, the authors would like to thank Ana Maria Semm, Renke Biallas, Jessica Feichtmayr, and Johannes Kissel, who assisted in conducting the study and analyzing the data, and Keri Hartman for proofreading. Finally, the first author (M Fink) would like to thank Larissa Kaltefleiter for her advice. This work was funded by the German Research Association (Deutsche Forschungsgemeinschaft; project number FOR2385).

Authors' Contributions

M Fink wrote the first draft of the manuscript, took part in conducting the study, and conducted data analysis and visualization. VR took part in conducting the study and provided feedback and editing. M Stadler conducted data analysis and visualization and provided feedback and assisted with editing. M Siebeck conceptualized and designed the study, provided feedback and editing, and acquired funding. FF conceptualized and designed the study, provided feedback and editing, and acquired funding. M Fischer conceptualized and designed the study, provided feedback and editing, and acquired funding. All authors approved the final manuscript for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participant characteristics across all conditions and CONSORT (Consolidated Standards of Reporting Trials)–style diagram of participant flow.
[DOCX File , 55 KB - jmir_v23i3e21196_app1.docx]

Multimedia Appendix 2
Overview of the experimental procedure and simulation phases.
[DOCX File , 22 KB - jmir_v23i3e21196_app2.docx]

Multimedia Appendix 3
Table containing the questions provided with all virtual patients. These questions were allocated to the five history-taking categories of main symptoms, prior history, allergies and medication, social and family history, and system review.
[DOCX File , 27 KB - jmir_v23i3e21196_app3.docx]

Multimedia Appendix 4
Authenticity scales, cognitive load scales, coding scheme for diagnostic accuracy, coding scheme for the quality of evidence generation, motivation scales, and details of the diagnostic knowledge tests.
[DOCX File , 33 KB - jmir_v23i3e21196_app4.docx]

Multimedia Appendix 5
Boxplots and bee swarm plots for authenticity, cognitive load, and clinical reasoning variables for standardized patients and virtual patients.
[DOCX File , 73 KB - jmir_v23i3e21196_app5.docx]

References

1. Edelstein RA, Reid HM, Usatine R, Wilkes MS. A comparative study of measures to evaluate medical students' performance. Acad Med 2000 Aug;75(8):825-833. [doi: 10.1097/00001888-200008000-00016] [Medline: 10965862]
2. Barrows HS, Abrahamson S. The programmed patient: a technique for appraising student performance in clinical neurology. J Med Educ 1964 Aug;39:802-805. [Medline: 14180699]
3. Botezatu M, Hult H, Tessma MK, Fors UGH. Virtual patient simulation for learning and assessment: superior results in comparison with regular course exams. Med Teach 2010;32(10):845-850. [doi: 10.3109/01421591003695287] [Medline: 20854161]
4. Vu NV, Barrows HS. Use of standardized patients in clinical assessments: recent developments and measurement findings. Educational Researcher 2016 Jul;23(3):23-30. [doi: 10.3102/0013189x023003023]
5. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J 1975 Mar 22;1(5955):447-451 [FREE Full text] [doi: 10.1136/bmj.1.5955.447] [Medline: 1115966]

6. Ziv A. Simulators and simulation-based medical education. In: A practical guide for medical teachers. Vol. 2. 3rd ed. Amsterdam: Elsevier; 2009.
7. de JT. Instruction based on computer simulations. In: Handbook of research on learning and instruction. New York: Routledge; 2011:446-466.
8. Villaume WA, Berger BA, Barker BN. Learning motivational interviewing: scripting a virtual patient. Am J Pharm Educ 2006 Apr 15;70(2):33 [FREE Full text] [doi: 10.5688/aj700233] [Medline: 17149413]
9. Setrakian J, Gauthier G, Bergeron L, Chamberland M, St-Onge C. Comparison of assessment by a virtual patient and by clinician-educators of medical students' history-taking skills: exploratory descriptive study. JMIR Med Educ 2020 Mar 12;6(1):14428 [FREE Full text] [doi: 10.2196/14428] [Medline: 32163036]
10. Heitzmann N, Seidel T, Hetmanek A, Wecker C, Fischer MR, Ufer S, et al. Facilitating diagnostic competences in simulations in higher education a framework and a research agenda. Frontline Learning Research 2019 Dec 3:1-24. [doi: 10.14786/flr.v7i4.384]
11. Kopp V, Stark R, Fischer MR. Fostering diagnostic knowledge through computer-supported, case-based worked examples: effects of erroneous examples and feedback. Med Educ 2008 Aug;42(8):823-829. [doi: 10.1111/j.1365-2923.2008.03122.x] [Medline: 18564096]
12. Chernikova O, Heitzmann N, Stadler M, Holzberger D, Seidel T, Fischer F. Simulation-based learning in higher education: a meta-analysis. Review of Educational Research 2020 Jun 15;90(4):499-541. [doi: 10.3102/0034654320933544]
13. Cook DA, Brydges R, Hamstra SJ, Zendejas B, Szostek JH, Wang AT, et al. Comparative effectiveness of technology-enhanced simulation versus other instructional methods: a systematic review and meta-analysis. Simul Healthc 2012 Oct;7(5):308-320. [doi: 10.1097/SIH.0b013e3182614f95] [Medline: 23032751]
14. Schubert T, Friedmann F, Regenbrecht H. The experience of presence: factor analytic insights. Presence: Teleoperators & Virtual Environments 2001 Jun;10(3):266-281. [doi: 10.1162/105474601300343603]
15. Shaffer DW, Resnick M. Thick authenticity: new media and authentic learning. J Interact Learn Res 1999;10(2):195-216 [FREE Full text]
16. Hamstra SJ, Brydges R, Hatala R, Zendejas B, Cook DA. Reconsidering fidelity in simulation-based training. Acad Med 2014 Mar;89(3):387-392 [FREE Full text] [doi: 10.1097/ACM.0000000000000130] [Medline: 24448038]
17. Hofer M. Presence und involvement. 1st ed. Baden-Baden: Nomos; 2016:978-973.
18. Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC. Acad Med 1993 Jun;68(6):443-451. [doi: 10.1097/00001888-199306000-00002] [Medline: 8507309]
19. Luctkar-Flude M, Wilson-Keates B, Larocque M. Evaluating high-fidelity human simulators and standardized patients in an undergraduate nursing health assessment course. Nurse Educ Today 2012 May;32(4):448-452. [doi: 10.1016/j.nedt.2011.04.011] [Medline: 21565436]
20. Rethans JJ, Sturmans F, Drop R, van der Vleuten C. Assessment of the performance of general practitioners by the use of standardized (simulated) patients. Br J Gen Pract 1991 Mar;41(344):97-99 [FREE Full text] [Medline: 2031767]
21. Friedman CP, France CL, Drossman DD. A randomized comparison of alternative formats for clinical simulations. Med Decis Making 1991;11(4):265-272. [doi: 10.1177/0272989X9101100404] [Medline: 1766329]
22. Padgett J, Cristancho S, Lingard L, Cherry R, Haji F. Engagement: what is it good for? The role of learner engagement in healthcare simulation contexts. Adv Health Sci Educ Theory Pract 2019 Oct;24(4):811-825. [doi: 10.1007/s10459-018-9865-7] [Medline: 30456474]
23. Norman G, Dore K, Grierson L. The minimal relationship between simulation fidelity and transfer of learning. Med Educ 2012 Jul;46(7):636-647. [doi: 10.1111/j.1365-2923.2012.04243.x] [Medline: 22616789]
24. Schoenherr JR, Hamstra SJ. Beyond fidelity: deconstructing the seductive simplicity of fidelity in simulator-based education in the health care professions. Simul Healthc 2017 Apr;12(2):117-123. [doi: 10.1097/SIH.0000000000000226] [Medline: 28704289]
25. La Rochelle JS, Durning SJ, Pangaro LN, Artino AR, van der Vleuten CPM, Schuwirth L. Authenticity of instruction and student performance: a prospective randomised trial. Med Educ 2011 Aug;45(8):807-817. [doi: 10.1111/j.1365-2923.2011.03994.x] [Medline: 21752077]
26. Sweller J, van Merrienboer JJG, Paas FGWC. Cognitive architecture and instructional design. Educational Psychology Review 1998;10(3):251-296. [doi: 10.1023/a:1022193728205]
27. Young JQ, Van Merrienboer J, Durning S, Ten Cate O. Cognitive Load Theory: implications for medical education: AMEE Guide No. 86. Med Teach 2014 May;36(5):371-384. [doi: 10.3109/0142159X.2014.889290] [Medline: 24593808]
28. Dankbaar MEW, Alsma J, Jansen EEH, van Merrienboer JJG, van Saase JLCM, Schuit SCE. An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. Adv Health Sci Educ Theory Pract 2016 Aug;21(3):505-521 [FREE Full text] [doi: 10.1007/s10459-015-9641-x] [Medline: 26433730]
29. Haji FA, Cheung JJH, Woods N, Regehr G, de Ribaupierre S, Dubrowski A. Thrive or overload? The effect of task complexity on novices' simulation-based learning. Med Educ 2016 Sep;50(9):955-968. [doi: 10.1111/medu.13086] [Medline: 27562895]
30. Young M, Thomas A, Lubarsky S, Ballard T, Gordon D, Gruppen LD, et al. Drawing boundaries: the difficulty in defining clinical reasoning. Acad Med 2018 Jul;93(7):990-995. [doi: 10.1097/ACM.0000000000002142] [Medline: 29369086]

31. Guagnano MT, Merlitti D, Manigrasso MR, Pace-Palitti V, Sensi S. New medical licensing examination using computer-based case simulations and standardized patients. Acad Med 2002 Jan;77(1):87-90. [doi: 10.1097/00001888-200201000-00020] [Medline: 11788331]
32. Hawkins R, MacKrell Gaglione M, LaDuca T, Leung C, Sample L, Gliva-McConvey G, et al. Assessment of patient management skills and clinical skills of practising doctors using computer-based case simulations and standardised patients. Med Educ 2004 Sep;38(9):958-968. [doi: 10.1111/j.1365-2929.2004.01907.x] [Medline: 15327677]
33. Hu L, Chen G, Li P, Huang J. Retracted article: multimedia effect in problem solving: a meta-analysis. Educ Psychol Rev 2019 Jul 11;32(3):901. [doi: 10.1007/s10648-019-09490-4]
34. Stillman PL, Swanson DB, Smee S, Stillman AE, Ebert TH, Emmel VS, et al. Assessing clinical skills of residents with standardized patients. Ann Intern Med 1986 Nov;105(5):762-771. [doi: 10.7326/0003-4819-105-5-762] [Medline: 3767153]
35. Woolliscroft JO, Calhoun JG, Billiu GA, Stross JK, MacDonald M, Templeton B. House officer interviewing techniques: impact on data elicitation and patient perceptions. J Gen Intern Med 1989;4(2):108-114. [doi: 10.1007/BF02602349] [Medline: 2709168]
36. Casus computer software. 2018. URL: https://www.instruct.eu/en/ [accessed 2021-01-30]
37. Bornemann BM. Documentation forms of internal medicine and surgery for history taking and the physical examination for the medical training of students in Germany: An analysis of content and structure. Diss. München: Institut für Didaktik und Ausbildungsforschung in der Medizin der Ludwig-Maximilians-Universität München; 2016. URL: https://edoc.ub.uni-muenchen.de/19166/1/Bornemann_Barbara.pdf [accessed 2021-02-16]
38. Seidel T, Stürmer K, Blomberg G, Kobarg M, Schwindt K. Teacher learning from analysis of videotaped classroom situations: does it make a difference whether teachers observe their own teaching or that of others? Teaching and Teacher Education 2011 Feb;27(2):259-267. [doi: 10.1016/j.tate.2010.08.009]
39. Vorderer P, Wirth W, Gouveia F, Biocca F, Saari T, Jäncke F, et al. MEC spatial presence questionnaire (MEC-SPQ): short documentation and instructions for application. Report to the European Community, Project Presence: MEC (IST--37661). 2001. URL: http://www.ijk.hmt-hannover.de/presence [accessed 2021-01-30]
40. Frank B. Validation. Measuring Presence in Laboratory-Based Research with Microworlds 2014:51-61. [doi: 10.1007/978-3-658-08148-5_6]
41. Opfermann M. There's more to it than instructional design: the role of individual learner characteristics for hypermedia learning. Berlin: Logos Verlag; 2008:1-295.
42. Rheinberg F, Vollmeyer R, Burns BD. FAM: Ein fragebogen zur erfassung aktueller motivation in lern- und leistungssituationen. Diagnostica 2001 Apr;47(2):57-66. [doi: 10.1026//0012-1924.47.2.57]
43. Wigfield A. Expectancy-value theory of achievement motivation: a developmental perspective. Educ Psychol Rev 1994 Mar;6(1):49-78. [doi: 10.1007/bf02209024]
44. Schmidmaier R, Eiber S, Ebersbach R, Schiller M, Hege I, Holzer M, et al. Learning the facts in medical school is not enough: which factors predict successful application of procedural knowledge in a laboratory setting? BMC Med Educ 2013 Mar 22;13(1):28 [FREE Full text] [doi: 10.1186/1472-6920-13-28] [Medline: 23433202]
45. Lakens D, Scheel AM, Isager PM. Equivalence testing for psychological research: a tutorial. Advances in Methods and Practices in Psychological Science 2018 Jun 01;1(2):259-269. [doi: 10.1177/2515245918770963]
46. Lakens D. Equivalence tests: a practical primer for tests, correlations, and meta-analyses. Soc Psychol Personal Sci 2017 May;8(4):355-362 [FREE Full text] [doi: 10.1177/1948550617697177] [Medline: 28736600]
47. R Foundation for statistical computing. R [Computer software]. Vienna, Austria: R Foundation for Statistical Computing; 2019. URL: https://www.r-project.org/ [accessed 2021-02-16]
48. Holm S. A simple sequentially rejective multiple test procedure. Scandinavian J Stat. 1979. URL: https://www.jstor.org/stable/4615733?seq=1 [accessed 2021-02-16]
49. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: Lawrence Erlbaum; 1988.
50. G*Power computer software. 2014. URL: https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html [accessed 2021-01-30]
51. Sweller J, van Merriënboer JJG, Paas F. Cognitive architecture and instructional design: 20 years later. Educ Psychol Rev 2019 Jan 22;31(2):261-292. [doi: 10.1007/s10648-019-09465-5]
52. Lapkin S, Levett-Jones T. A cost-utility analysis of medium vs. high-fidelity human patient simulation manikins in nursing education. J Clin Nurs 2011 Dec;20(23-24):3543-3552. [doi: 10.1111/j.1365-2702.2011.03843.x] [Medline: 21917033]
53. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990 Sep;65(9 Suppl):S63-S67. [doi: 10.1097/00001888-199009000-00045] [Medline: 2400509]
54. Lewis KL, Bohnert CA, Gammon WL, Hölzer H, Lyman L, Smith C, et al. The Association of Standardized Patient Educators (ASPE) Standards of Best Practice (SOBP). Adv Simul (Lond) 2017;2:10 [FREE Full text] [doi: 10.1186/s41077-017-0043-4] [Medline: 29450011]

Abbreviations

CG: case group

NHST: null hypothesis significance testing
SP: standardized patient
TOST: 2 separate 1-sided test
VP: virtual patient

Edited by G Eysenbach, R Kukafka; submitted 25.06.20; peer-reviewed by J Cheung, P Bergl, A Kononowicz, S Edelbring; comments to author 08.08.20; revised version received 01.10.20; accepted 27.12.20; published 04.03.21.

Please cite as:
Fink MC, Reitmeier V, Stadler M, Siebeck M, Fischer F, Fischer MR
Assessment of Diagnostic Competences With Standardized Patients Versus Virtual Patients: Experimental Study in the Context of History Taking
J Med Internet Res 2021;23(3):e21196
URL: https://www.jmir.org/2021/3/e21196
doi:10.2196/21196
PMID:33661122

Original Paper

Basic Medical Training for Refugees via Collaborative Blended Learning: Quasi-Experimental Design

Thibault Lovey[1*], MMed; Paul O'Keeffe[2*], PhD; Ianis Petignat[3*], MMed

[1]Department of Public Health & Global Health, Division of Infectious Diseases, Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland

[2]InZone, University of Geneva, Geneva, Switzerland

[3]Division of Infectious Diseases, Hôpitaux Universitaires de Genève, Geneva, Switzerland

[*]all authors contributed equally

Corresponding Author:
Paul O'Keeffe, PhD
InZone
University of Geneva
Boulevard Pont d'Arve 28
Geneva, 1205
Switzerland
Phone: 41 223799533
Email: Paul.okeeffe@unige.ch

Abstract

Background: Globally, there is an excess of 68.5 million people who have been forced to leave their homes and seek sanctuary elsewhere because of poverty, persecution, conflict, violence, and human rights violations. Although international humanitarian responses usually focus on ensuring that the basic needs of these people are being met, there is growing attention on the role that development-oriented interventions can play in the longer term. Higher education in a refugee context is one such intervention that can equip refugees with the knowledge and skills they need to serve their communities and move forward.

Objective: This study aims to evaluate the outcomes and effectiveness of the University of Geneva InZone-Raft Basic Medical Training Course in the Kakuma refugee camp in Kenya compared with a previous incarnation of the same course in the Dadaab refugee camp in Kenya.

Methods: We used a quasi-experimental design to compare the posttest scores of both inequivalent student groups: control group (n=18) and intervention group (n=16). Factors that influenced refugee students' knowledge acquisition, the amount of knowledge they acquired, and their academic outcomes were assessed, and the pedagogical evolution of the project is presented.

Results: We found that the Kakuma intervention course yielded better outcomes and was more effective in terms of learning than the Dadaab control course. Of the 16 students who took part in the intervention course, 10 (63%) completed the program successfully and received accreditation from the University of Geneva. We observed that they received new knowledge well and scored higher on all learning modalities than those in the control course. Comparison of written and oral examinations between the courses showed statistical significance for the intervention group in written and oral exams (two-tailed: P=.006 and P=.05; one-tailed: P=.003 and P=.03, respectively). The Kakuma course was not effective in addressing electricity and internet access problems, nor in reducing the challenge of tight deadlines in the syllabus. Pedagogical adjustments to the intervention course improved student involvement, with higher participation rates in quizzes (10/11, 91%), and overall satisfaction and learning.

Conclusions: The intervention group—with an improved mode of delivery, better contextualized content, and further interaction—reached a higher level of medical knowledge acquisition and developed more complex questions on medical topics than the control group. The positive outcome of this project shows that given the right resources and support, refugees can contribute to the improvement and development of health care in their communities. Nonetheless, a more focused effort is necessary to meet the educational needs of refugee learners and better understand their living conditions.

(J Med Internet Res 2021;23(3):e22345) doi:10.2196/22345

KEYWORDS

refugees; blended learning; basic medical training; higher education in emergencies; innovation; mobile phone

Introduction

Higher Education in Refugee Contexts

Academic discourse on the role of higher education has long centered on liberatory or functional objectives seen through humanistic or utilitarian lenses. Regardless of one's ideological stance on these matters, most academics agree that the mission of higher education is multifaceted [1] and hinges on 3 important goals: delivering education, generating new knowledge, and engaging with society [2]. Through established pedagogical systems and social institutions, human progress has harnessed education to clear a pathway for those who participate within its confines to better understand, learn, and succeed.

Although all 3 goals of higher education play a critical role in constructing and understanding the world around us, the *society engagement* role is the engine in which we instrumentalize, form, govern, interact, and negotiate with the society or community and the wider world in which we exist. Whether economic, political, or social, societies in which higher education and social engagement are a priority are generally held to be successful and desirable places to live in [3]. In societies in which they are absent or lacking, there are many negative social implications, not just affecting theoretical future human progress but also real life here and now.

Emergencies, fragile states, poverty-stricken communities and refugee contexts are all places where higher education is absent, lacking, underresourced, and underresearched. Higher education in refugee contexts is a relatively new academic field of pedagogical scholarship that focuses on the specifics and outcomes of higher education in refugee communities. Growing out of sustained global efforts in the latter half of the 20th century and the early part of the 21st century to use primary and secondary level education as a driver for development, higher education in refugee contexts has been primarily rooted within the global forced migration management system as it attempts to contend with ever-growing populations of displaced and desperate people. As an increasing number of people enter the global forced migration system and/or come up through its management infrastructure, higher education and its application for human development is gaining traction as a key solution that may provide hope for averting or deepening humanitarian crises. Global initiatives such as the Sustainable Development Goals (SDGs) for 2030 have firmly positioned higher education as a policy direction [4]; however, little attention has been paid to developing appropriate pedagogical models that work for most of the globally displaced—refugees stuck in underresourced and overburdened refugee camps in poor and fragile states.

The Reality of the Globally Disposed

The global refugee and displaced population is currently in excess of 68.5 million people, with one person forcibly displaced every second [5]. Most (85%) of the globally displaced live in low- and middle-income countries, with little access to basic services such as education and health care [6]. Although this unprecedented number grows daily, the very nature of the conflicts and abuses driving these figures often results in refugee populations spending excessive amounts of time—5 years or more—in exile because of the protracted nature of the situations in which they find themselves [7]. According to the United Nations High Commissioner for Refugees (UNHCR), nearly 16 million people (78% of the total world refugee population) found themselves in protracted refugee situations in 2019—a 12% increase in figures from the previous year [8]. Through no fault of their own, more young people are finding themselves stuck in limbo-like situations in refugee camps close to conflict zones for longer. Their dwindling life prospects, economic disempowerment, vulnerability to health crises, and many other social and political ills have given them less of a chance to live the kind of lives they frequently glimpse at through foreign movies and websites on their smartphones.

Recognizing the limiting life opportunities offered by refugee camps, the desire of young people to advance their own lives, and indeed the necessity of finding durable solutions to the disruption that displacement brings, humanitarian responses have begun to seek more sustainable approaches. High on the international agenda is the move beyond immediate humanitarian interventions toward development as a goal in itself. The 2030 Agenda for Sustainable Development, encapsulated by 17 SDGs, or a "blueprint for peace and prosperity," enunciated this new direction by stating that the recognition "that ending poverty and other deprivations must go hand-in-hand with strategies that improve health and education, reduce inequality, and spur economic growth" [9].

Development Through Education or the Status Quo?

With greater emphasis being placed on role development in finding durable solutions to displacement problems, the education for development conversation has moved on from the provision of primary and secondary education to include the role that higher education can play in development. The call for "lifelong learning opportunities for all" in SDG 4 [9] has further added power to the argument that higher education can provide a development solution for displaced populations.

Furthermore, higher education lies at the heart of the resilience and empowerment discourse popular in forced migration academia, policies, and practice. Education is often touted as providing a sense of purpose amid the uprootedness of refugee status and life constraints in refugee camps [10,11]. It is championed for playing a vital role in facilitating endurance and transitions for refugees by providing them with the skills they need to increase their social capital and ability to adapt to different and challenging contexts [12-15].

However, despite all the praise, the academic inquiry into the role of higher education in refugee contexts is not always positive. Critical academic approaches that take a more liberatory focus question the impact of higher education in propagating a system that keeps refugees contained and oppressed. In addition, the evolution of higher education in refugee contexts is often characterized as a neocolonial enterprise. Critics point to most of the teaching and learning programs accessible to refugees coming from Western institutions [16], being noncontextualized for refugee contexts and delivered through Western perspectives [17] via Western-centric web-based education platforms [18].

Accessing Higher Education

Although there may be many questions left to answer about the purpose of higher education in refugee contexts and its intersection with authority, power, and control, there is the certainty that the pedagogical conditions that students in refugee camps have to contend with do not match with the condition of students elsewhere [19]. Some of the major obstacles they face in participating in higher education programs include the physical space where they find themselves, security issues that emanate from states of fragility, and major resource deficits that they experience [20]. Traditional *brick and mortar* higher education institutions tend to be in urban areas and/or require infrastructure that is not conducive to refugee camps and are often out of reach of refugees. Advancements in communication technology and web-based learning platforms have begun to address these shortcomings by paving the way for other possible solutions for bringing higher education to remote and inhospitable refugee camps. Massive open online courses (MOOCs; although problematic in themselves as they require technical capacity and tend to be designed with nonrefugee audiences in mind) and other web-based platforms—provided there is a technical capacity in place—can offer learners in refugee camps higher education opportunities they otherwise would not have. However, opportunities aside, to be successful, refugee students often require fortified scaffolding to support their learning [20].

In a synthesis of the literature, Bawa has found that web-based learning programs have a dropout rate of 40% to 80% in high resource contexts [21]. In situations such as refugee camps, where electricity is not always guaranteed, and computers and the internet are rare and expensive commodities, web-based learning may be a less than appropriate option [19]. On the other hand, blended learning approaches may offer the necessary support to ensure success for refugee learners. Although research and the literature on the benefits of blended learning in refugee camps are hard to come by, a pedagogical model, which has been found to be successful (and central to this case study), is the InZone collaborative blended learning model. This model has been found to achieve a course completion rate of 82.6% for refugee students who participated in higher education courses in Kakuma and Azraq refugee camps between 2017 and 2018 [22]. The model is presented below.

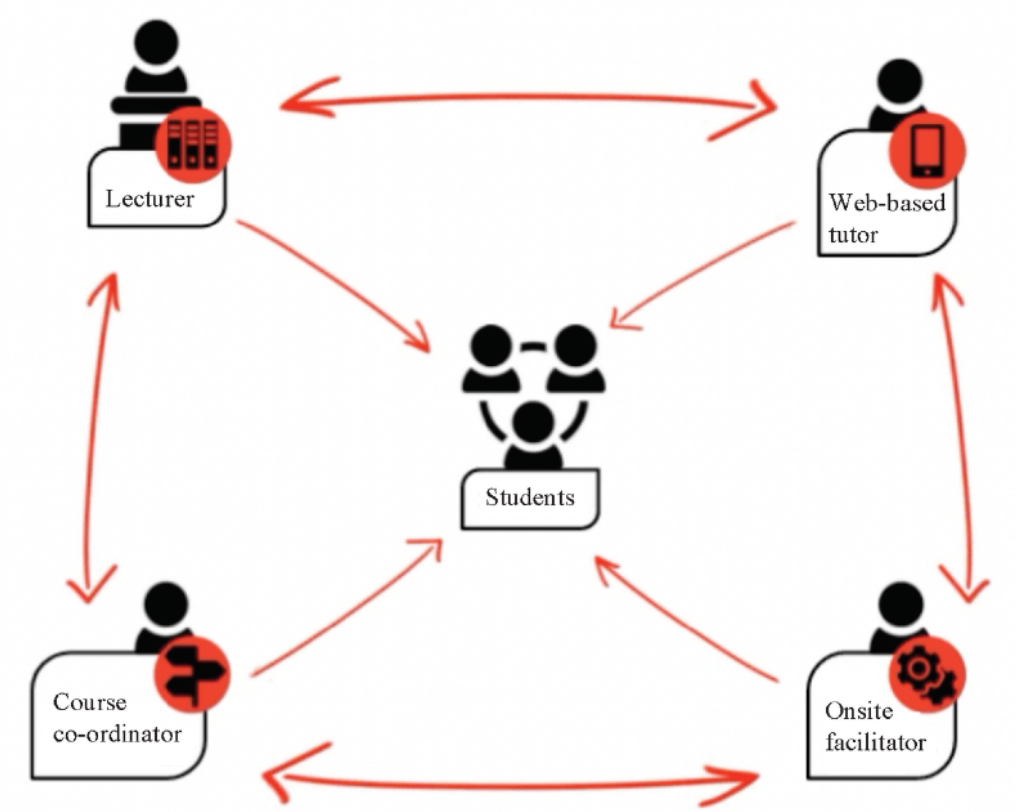
Kakuma Refugee Camp

Located in the remote and arid Turkana state in North-Western Kenya, the Kakuma refugee camp is home to over 197,341 refugees from around 20 different nationalities, representing 39% of all registered refugees in Kenya [23]. Nationalities include South Sudanese, Somalians, Congolese, Ethiopians, Burundians, Sudanese, Ugandans, Rwandans, and Eritreans [24]. The camp was created in 1992 in response to the civil war in Sudan, which saw 20,000 young boys (the *lost boys of Sudan*, seeking sanctuary in neighboring countries [24]). Climate conditions are extreme. Temperatures routinely reach 38° C during the hot season, and rainfall is sparse. The camp is administered by the UNHCR and the Kenyan state and follows a containment policy with strict restrictions on movement in, out of, and around the camp. It is divided into 2 operative areas: Kakuma refugee camp and Kalobeyei integrated settlement. Although relatively safe compared with other refugee camps in the region, there are frequent security breaches [25]. Health care in Kakuma is serviced by 4 small clinics and Kakuma Mission Hospital, which has 62 beds [26]; however, overcrowding, inadequate sanitation, and poor access to these basic health services make Kakuma a dangerous place to live [27].

InZone-Raft Basic Medical Training Course

InZone, an academic and humanitarian program at the University of Geneva, has been active in the Kakuma refugee camp for a decade, where it enables accredited higher education courses via a collaborative blended learning ecosystem. InZone has a center (the learning hub) in Kakuma where students come to access computers and Wi-Fi, meet with each other, and connect to web-based courses from the University of Geneva. The InZone learning ecosystem, which scaffolds the students, takes a student-centered approach where learning materials and lectures are delivered through the web and facilitated by a refugee management team in a refugee camp. Students receive pedagogical support from web-based tutors from the University of Geneva, who *meet* the students through online tutorials, via WhatsApp instant messaging, for a designated amount of time each week throughout the course. Students are also supported by trained onsite facilitators in the learning hub, who assist the tutors in managing the class and the learning process. The learning ecosystem is overseen by a course co-ordinator who helps ensure the overall smooth running of the course. Figure 1 depicts the InZone collaborative learning ecosystem, with the following explaining the role of each of its actors.

Figure 1. InZone collaborative learning ecosystem.



Key Actor Roles and Responsibilities in the Collaborative Learning Ecosystem

1. The lecturer delivers the course material over a learning platform (eg, via MOOCs freely available on Coursera or EdX), encourages the generation of new knowledge, and evaluates the students' learning. In the ecosystem, the delivery of knowledge via a web-based platform enables the transmission of information to the students, who through discussions, group work, and so on acquire and develop new knowledge.
2. The web-based tutor is a subject matter expert or peer with a more advanced level of subject knowledge. The tutor plays a pedagogical role in this collaborative learning ecosystem by *meeting* the students regularly over an information communication technology platform (eg, WhatsApp) to stimulate new knowledge acquisition, discuss the student's progress, and offer advice on being a successful learner. The web-based tutor also travels to the camp to meet the students in person and deliver face-to-face classes toward the end of the course.
3. The onsite facilitator provides onsite technical and guiding support to learners, helping them access the learning platform on location and navigate the physical learning space. The onsite facilitator is a critical contact point in the educational relationship among students and other members of the collaborative learning ecosystem as they are in frequent physical contact with the students.
4. The course co-ordinator has the overall responsibility for the day-to-day running of the course and liaises with other members of the learning ecosystem to ensure a smooth operation.
5. The student is the focal point of the learning ecosystem. This means that they are central to the collaborative learning model, and the entire learning ecosystem is designed to support their optimal learning by meeting their educational needs and promoting progressive learner autonomy [22].

In 2017, following the successful implementation of a basic medical training course in the Dadaab refugee camp in Northern Kenya, InZone was approached by Kakuma refugee camp authorities to develop a health care course for residents in the camp. The impetus for developing the course was the hope that locally trained refugees could partially fill the lack of staff in the health care sector in the camp [28]. To do this, InZone connected with the Raft telemedicine network at the University of Geneva and created a basic medical training course that could be enabled through InZone's collaborative learning ecosystem in the camp.

Using lessons learned from a similar medical education program in Dadaab, the basic medical training course aimed to contextualize the curriculum, use a pedagogical methodology that best suits students' needs, and ultimately prepare refugees to integrate into the health care services in the camp.

Research on Higher Education in Refugee Contexts

As mentioned earlier, higher education in refugee contexts is an emerging field of study, with little literature covering medical pedagogy in refugee camps. To the best of the authors' knowledge, little research has been carried out into which pedagogies would best satisfy the needs of refugee learners and the impact medical higher education can have on their communities. This study aims to add empirical evidence to this

emerging field of study by evaluating the InZone-Raft basic medical training course across the 2 locations.

Research Objectives

The primary objective of this research project is to compare the grades obtained by students in the Kakuma intervention course with the grades of the students in the Dadaab control course. This objective evaluates whether the students following a newer educational method employed in Kakuma would have higher scores in learning modalities (written and oral) than the control group in Dadaab, who followed an older educational model.

The second objective of this research is to evaluate the effectiveness of the new education model by quantifying the amount of knowledge acquired in the 2 groups. It was postulated that the Kakuma group would show an improvement in their acquisition of basic knowledge throughout the course compared with the control group in Dadaab.

The third objective of this research is to identify the confounding variables that influenced the final grades and thus the acquisition of new knowledge. It was assumed that external factors impact the acquisition of knowledge, particularly in a crisis context, and that these factors are common to the Kakuma and Dadaab cohorts.

Research Questions

1. Did students who applied the newer contextualized blended learning methodology in the Kakuma intervention course obtain better posttest results than students who applied the traditional methodology in the Dadaab control course?
2. Was the education model used in the intervention course more effective in producing better periodical knowledge acquisition at set points throughout the course than the education model used in the control course?
3. What confounding variables may have influenced the student's posttest results in both groups?

Methods

Study Design

Owing to ethical considerations, it was not possible to divide a cohort of refugee students for the purpose of this research. Consequently, randomization was not possible, and the study was conducted as a quasi-experimental study. The control group was nonequivalent and recruited from another classroom that was already following the older teaching model at the time this study began.

Refugee camps challenged the practicality of the study, as certain external factors could not be controlled. There are numerous barriers related to access to education that are present in refugee contexts and are therefore unavoidable. In this specific environment, the quasi-experimental design has the advantage of providing higher external validity by allowing the experiment to be implemented directly without the need for artificial modifications.

Inclusion Criteria and Sample Size

Participants were recruited over 2 months, from August to September 2018, in the Kakuma refugee camp through the InZone selection process. To be admitted to the course, the applicant had to be a resident in the Kakuma refugee camp or Kalobeyei integrated settlement (a nearby refugee–host community settlement), own a high school diploma, satisfy the proficiency requirements for the English language with a C grade in the Kenyan high school state exam or equivalent, and write a motivation essay (Multimedia Appendix 1).

Functions for power analysis were used to evaluate the appropriate size of the intervention group [29]. The parameters were set to a significance level of .05, a power of .8, with a control group of 18 participants and an expected effect size of 1.2. The formula estimated the minimum sample size required for the two-tailed and one-tailed *t* tests of 9 and 6 students, respectively. With an estimated dropout rate similar to that of the control group, at 33% (9/27), the recruitment was set with a total of 16 students in the intervention group.

Intervention

The basic medical training course consisted of 3 modules. Module 1 consists of an introduction to the major organ systems of the human body. It covers the physiology and anatomy of 13 organ systems. Module 2 is more clinically oriented and teaches basic illness physiopathology relevant to medical conditions in sub-Saharan Africa. Module 3 is a case-based learning unit in which students investigate the medical conditions of 9 patients. The content of the modules was discussed and evaluated by a multidisciplinary professional team composed of physicians, a pedagogue, and a sociologist to ensure high quality of content relevant to the refugee situation and consistent with the level of a higher education course.

To avoid the additional costs and difficulties of sending teaching materials to refugees and ensure almost instantaneous deployment of the course, all materials were available on the web at the University of Geneva's Moodle learning platform (Moodle Pty Ltd) [30] in which the students had access to 13 units divided by organ system. Each unit was built identically with the corresponding Khan Academy videos (Khan Academy Ltd) [31], chapters from the OpenStax e-book *Anatomy and Physiology* [32], various additional documents to better strengthen their understanding of the topic, and a list of key concepts to be mastered for their final evaluation (Multimedia Appendix 2).

Computers and free Wi-Fi were accessible in the learning hub for the students to access at designated time slots throughout the course to ensure equitable accessibility among all course participants. Transport and meals at the hub were subsidized by course funds for the same reasons. Huawei Media Pad T3 10 tablets were also allocated to the students so that they could access preloaded learning materials offline outside of the allocated access times at the learning hub. A student was appointed as an onsite facilitator to assist in managing the day-to-day activities of the class and assist web-based tutors with lesson co-ordination.

At designated times each week, 2 medical students from the University of Geneva tutored the students via WhatsApp using structured web-based tutorials. These tutorials involved the presentation of questions and discussion points on a WhatsApp

forum at the start of each week. Students and tutors later synchronously and asynchronously discussed these points and questions throughout the week. On average, for module 1, one system was studied each week with planning adjusted according to any difficulties experienced by the students (eg, flooding hindering students' travel to the learning hub).

The procedure for the control group in the Dadaab refugee camp was identical except that tablets and Khan Academy videos were not provided to students.

Sampling

The study received 45 applications to take part in the new education model in the Kakuma refugee camp. Following a selection procedure carried out by InZone, a cohort of 16 students was recruited to participate in the InZone-Raft basic medical training course, which ran from October 2018 to February 2020. They took part in module 1 of the course from October 2018 to February 2019.

Out of the 16 students who initially started the course, 12 completed it, 11 took a final written and oral exam, and 10 passed it. The reasons for dropping out of the course were *family problems* (n=2), moving out of the refugee camp (n=1), attending an alternative training course (n=1), and lack of commitment because of relocation (n=1). A total of 11 students were included in the intervention group for the statistical analysis, as they presented themselves in the final oral and written exam (Multimedia Appendix 3).

The control group consisted of 18 students from the cohort who followed the previous education model from February 2017 to February 2018 in the Dadaab refugee camp. Of the 27 students initially recruited, 18 completed the course and were included in the statistical analysis. Reasons for dropping out included relocation of non-Somali refugees to Kakuma camp, return to Somalia, relocation to other countries, and family problems [33] (Multimedia Appendix 3).

Demographic data (age, sex, and country of origin) and educational attainment data were collected from all participants during the application process.

Measures and Instruments

The objectives of this quasi-experimental study focus on 3 particular aspects of module 1 that need to be measured in both groups: (1) posttest results, (2) knowledge acquisition, and (3) confounding variables.

Posttest Results

To ensure adequate measurement of student performance at the end of the first module, tutors traveled to the respective refugee camp to provide face-to-face revision classes for 2 days and a day of summative evaluation, including a supervised written exam and an oral exam. The written examination consisted of 53 multiple-choice questions and 18 development questions. The oral examination consisted of 6 questions related to the organ system described in a clinical situation and 2 additional questions related to a different system. The content of the written and oral examinations was discussed with the multidisciplinary team to ensure it was representative of all the topics covered in module 1, thus guaranteeing an adequate balance of questions.

Several methods have been used to ensure the highest possible reliability in the measurement of student knowledge. The written examination used multiple-choice questions to ensure impartiality and was marked anonymously by 2 different experts, and the marks were then cross-checked. The oral examinations used a standardized assessment grid, and student performance was assessed by 3 experts who were physically present at the assessment (written exam: Cronbach α=.92, 95% CI 0.86-0.98; oral exam: Cronbach α=.92, 95% CI 0.86-0.98) [34].

Quantitative data on written and oral examination results, final grades, and critical thinking questions were collected directly from the respective teachers for both cohorts at the end of module 1.

Knowledge Acquisition

Quizzes were used to assess the students' progression at regular intervals throughout the course (at the end of each unit of study). These formative evaluations were divided into 3 levels of difficulty and could be repeated, if desired; a multidisciplinary team designed them to ensure that it assesses the acquisition of the key elements of each organ system addressed.

To assess student satisfaction and give suggestions for improvements, a web-based feedback survey was disseminated to the students at the end of each study unit and at the end of each module.

The quiz results and feedback for each unit were extracted from Moodle in the csv (comma-separated values) format. The module 1 course feedback was sent after the final exams to the students who completed the course via Google forms and then exported to Microsoft Excel.

Confounding Variables

To ensure sufficient internal validity, variables that could influence posttest results such as average quiz scores, Moodle logs, number of WhatsApp messages, time used to complete the written examination, number of hours of study per week, previous medical experience, and gender were taken into account in the analysis and interpretation of total scores.

For the activity on the students' Moodle platform, the logs were extracted from Moodle for each student. The WhatsApp discussions between students were extracted directly from the application as a text file. The collection of Moodle logs, WhatsApp discussions, and critical thinking questions could only be obtained from the Kakuma cohort for technical reasons.

Statistical Analysis

To protect the anonymity of the students, all names were coded before performing the statistical analysis. All statistical tests used an α of .05 as the significance level. R Studio [35] was used as the main part of the analysis.

Posttest Results

A descriptive analysis was carried out for the results of the written and oral examinations, the final score, and the quizzes of module 1 [34] to evaluate the results of both groups. The

mean and median were used as central indicators of dispersion and expressed with a CI and a minimum and maximum value to assess the range.

To demonstrate whether the students who participated in the new method performed better than the control students, a posttest analysis was performed to assess the impact of the intervention, comparing the average between the 2 groups for the written exam, the oral exam, the final score, and the specific quiz questions with the overall statistics [36]. A one-tailed and two-tailed Student *t* test was used, and the nonparametric Wilcoxon signed-rank test was used when the distribution of scores was not normal.

Knowledge Acquisition

Only the first attempt was taken into account for the individual calculation of the average of each quiz. In addition, before the calculation of the individual average in the quiz of module 1, the 2 most extreme values of the individual quizzes were removed [37].

Spearman correlations were used to look for patterns between the quizzes and the final score to determine whether a method was more effective in producing knowledge acquisition throughout the course.

Confounding Variables

A WhatsApp discussion analysis [38] and text mining and sentiment analysis on WhatsApp text were performed to assess confounding variables that may influence student scores [39-42]. The student feedback was also evaluated [43].

Multiple regression analyses—with robust methods and variable transformations to ensure normality—were used to assess any confounding variables (including average results of the quizzes, Moodle logs, the number of WhatsApp messages, the time they used to complete the written exam, the number of hours per week they studied, their previous experience in medicine, and their gender) that might influence the final score in both groups [44-48].

Ethical Considerations

This study was approved by the Swiss Ethics Committee on research involving humans. The protocol number assigned was 2017-00632. Participants were informed of the purpose of the research, the methods used, and that enrollment, successful completion, or withdrawal from this program did not affect any eventual resettlement or repatriation process.

Participation was voluntary, and all the participants provided written informed consent. The *WhatsApp Forum Guidelines for InZone-RAFT Basic Medical Training* code of conduct was also accepted and signed by all students before using the WhatsApp group (Multimedia Appendix 1). This study included qualitative data from a small sample of a vulnerable population. These data cannot be released to the public because of the potential risk of identifying vulnerable refugees. Special requests for access to the data were evaluated in collaboration with the Ethics Commission of the University of Geneva.

Results

Sociodemographic Characteristics

In the control and intervention groups, women accounted for 27% (control: 5/18; intervention: 3/11) of the total students. The country-of-origin data varied in both groups. In the intervention group (Kakuma), the distribution of country of origin was diverse among the students, with 18% (2/11) from Burundi, 27% (3/11) from the Democratic Republic of Congo, 27% (3/11) from Southern Sudan, 9% (1/11) from Sudan, 9% (1/11) from Ethiopia, and 9% (1/11) from Somalia. In the control group, the representation was narrower, with 94% (17/18) from Somalia and 6% (1/18) from the Democratic Republic of Congo. Prior knowledge or experience in health-related areas was present in 36% (4/11) of the intervention group and 39% (7/18) of the control group. Age distribution could not be calculated as, for sociocultural and political reasons, dates of birth are not always recorded by families or they are artificially generated by the asylum-granting authorities when individuals apply for refugee status (Multimedia Appendix 3).

The homogeneity test concerning the sociodemographic characteristics of the 2 groups showed no significant differences and therefore were both considered homogeneous.

Posttest Results

For the written exam, with a possible total of 75 points, we obtained a mean score of 38 (SD 14; 51%) points, a median score of 35 (46%) points with a maximum score of 60 (80%) points, and a minimum score of 13 (17%) points. A question in the exam reached a perfect score (Q19), and 4 questions (Q28, Q33, Q43, and Q45) obtained a percentage of correct answers lower than 20% (2/11). For the oral exam, with a possible total of 20 points, we obtained a mean score of 13 (SD 5.5; 67%) points, a median score of 14 (70%) points with 1 perfect score, and a lower score at 2 (10%) points, which represents a question answered correctly. For critical thinking questions, with a possible total of 8 points, we obtained a mean score of 4 (SD 2.6; 53%) points, a median score of 4 (50%) points, 1 perfect score, and a minimum score of 1 (13%) point. If we combined the oral and written examinations by weighting the results by half between oral and written examinations, the mean percentage was 59% (SD 21.9) and the median percentage was 56% (range 14%-88%). For the Dadaab cohort of students studying the same course (control group), the mean percentage was 41%, median was 40%, maximum grade was 71%, and minimum grade was 15% (absolute numbers were not available; Table 1). We noted a significant difference between the previous cohort and this cohort in the total score (two-tailed: P=.008; one-tailed: P=.004). The comparison between the Dadaab and the Kakuma cohorts' written and oral examinations showed statistical significance for the intervention group in the written and oral exams (two-tailed: P=.006 and P=.05; one-tailed: P=.003 and P=.03, respectively), as shown in Table 1. Comparing the 5 identical developmental questions on the written test between the Dadaab and Kakuma cohorts yields a significant difference in means for this year group (two-tailed: P=.03; one-tailed: P=.02).

Table 1. Results by evaluation type for intervention and control groups.

Evaluation type	Mean			*P* value		Median			Range (%)
	Mean score (SD)	Percentage	95% CI	Two-tailed	One-tailed	Median score	Percentage	95% CI	
Experimental group									
Written exam (total possible score: 75 points)	38 (14)	51	38-63	.006[a]	.003[a]	35	46	36-75	17-80
Oral exam (total possible score: 20 points)	13 (5.5)	67	46-56	.05[b]	.03[b]	14	70	40-95	10-100
Final results	—[c] (21.9	59[d]	44-74	.008[e]	.004[e]	—[c]	56[d]	43-88	14-88
Critical thinking (total possible score 8 points)	4 (2.6)	53	32-75	N/A[f]	N/A	4	50	13-88	13-100
Quiz results (total possible score: 10 points)	5 (1.3)	48	40-56	N/A	N/A	5	46	38-66	29-69
Control group									
Written exam	—[g]	30	23-38	N/A	N/A	—[g]	31	22-39	0-56
Oral exam	—[g]	51	43-59	N/A	N/A	—[g]	46	40-53	30-93
Final results	—[g]	41	34-48	N/A	N/A	—[g]	40	36-45	15-71

[a]Comparison of written exam mean scores between groups (Student *t* test).

[b]Comparison of oral exam means between groups (Wilcoxon signed-rank test).

[c]Not available.

[d]Calculated by combining written and oral examination results weighted by half.

[e]Comparison of final results mean scores between groups (Wilcoxon signed-rank test).

[f]N/A: not applicable.

[g]Absolute numbers not provided by teachers.

Knowledge Acquisition

The intervention cohort's average for the 19 quizzes for module 1, with a possible total of 10 points each, had a mean score of 5 (48%) points, a median score of 5 (46%) points, a maximum score of 7 (69%) points, and a minimum score of 3 points (29%; Table 1). The median participation rate was 91% (10/11) of students for the cohort. There was a positive correlation between the quizzes and written scores (ρ=0.93; *P*<.001) and an even stronger correlation with the total score (ρ=0.94; *P*<.001). A total of 25 quiz questions used for the control cohort were reused for the intervention cohort, and it was found that 4 quiz questions had a statistically significant higher score for last year's group and 2 for this year's group.

Confounding Variables

WhatsApp Forum

On the WhatsApp discussion forum, which comprised 21 members, the chat started on October 22, 2018, at 9:20 AM. In total, 677 messages were sent to the participants (including 148 media messages). The most active day on the forum was October 28, with 23 messages sent, the most popular day for sending messages was Saturday (n=112), and the preferred time for texting was 11 AM (n=54; Multimedia Appendix 4).

The most active students were S10 (n=204), followed by S5 (n=57) and S12 (n=44), whereas those who answered the tutors' questions the most were S10 (n=44), followed by S12 (n=12) and S5 (n=9). Sentiment analysis of 10 different feelings was performed using text mining techniques, and the most representative sentiment was *positive* (n=516), followed by *trust* (n=344), and the least representative was *disgust* (n=39; Multimedia Appendix 5).

Feedback

In the feedback form, we found that the main source of learning materials was the Khan Academy videos (n=11), followed by the WhatsApp group forum (n=4).

Students accessed the InZone learning hub 5 times per week (median) and studied 20 hours (median). In general, the feedback was positive for the 11 students who finished module 1. The level of the content (8/11, 73%), amount of information to learn (8/11, 73%), and the level of English (9/11, 82%) were all adequate for the students. An exception was that the students felt the time available to study was too short for 7 (64%) of them. The main obstacles highlighted for the course were access to electricity (8/11, 73%) *every day*, internet access (7/11, 64%) *every day*, and access/transport to the hub (6/11, 55%) *every day*. Cultural barriers (10/11, 91%) *never*, lack of teaching

support (6/11, 55%) *never*, lack of prior education (5/11, 45%) *never*, lack of space to study (8/11, 73%) *never*, and money (7/11, 64%) *once a month* were not identified as potential barriers. Personal issues were more variable among the students *rarely* (mode).

Multiple Linear Regression

Multiple linear regression was calculated to predict the final score in the intervention group based on the average of the quizzes, Moodle logs, the number of WhatsApp messages, the time they used to complete the written exam, the number of hours per week they studied, their previous experience in medicine, and their gender. A significant regression model was found ($F_{3,7}$=23.89; P<.001), with an R^2 adjusted to 0.87. The participants' predicted final score was equal to 1.01−0.19 (exam time)+0.28 (previous experience)−0.40(gender), where exam time was measured in hours, previous experience was coded as 1=yes, 0=no, and gender as 1=woman, 0=man (Table 2). The participants' final score diminished by 19% for each extra hour, previous knowledge increased the final score by 30%, and women scored on average 40% less than men if all other variables remained constant. Regression in control groups holds the same conclusion and therefore is not detailed.

Table 2. Multiple linear regression to predict the final score in the intervention group.

Predictors	Estimates	95% CI	*P* value
Intercept	1	0.68 to 1.33	<.001
Written exam time	−0.19	−0.33 to −0.04	.02
Previous experience	0.28	0.15 to 0.40	.001
Gender	−0.4	−0.54 to −0.27	<.001
Observations	11	N/A[a]	N/A
R^2	.91	0.84 to 0.98	N/A
R^2 adjusted	.87	0.77 to 0.97	N/A

[a]N/A: not applicable.

Discussion

Posttest Comparison Between the Cohorts

To evaluate the Kakuma students' learning throughout the course, we started by looking at their exams results. We grouped the cohort into 3 main groups: a group with 3 students who exceeded our expectations, a pre-eminent group that met the minimum conditions, and a student who was behind the others. The results follow a Gaussian distribution and reflect the teachers' initial assessment of the student's level. In addition, teachers were impressed by the complexity of the questions asked by the students during revision sessions. In their opinion, highly motivated students attained a level of medical knowledge comparable with the first-year class in a medical school in the West.

One of the main objectives of this study was to evaluate the effectiveness of the new education model. If we consider the average score for all students in both cohorts, we have a significant difference in both two-tailed and one-tailed results for the written exam, oral exam, and total score. Overall, we can say that the Kakuma cohort had a better overall grade, but a disparate comparison of exams is questionable. We cannot exclude that this intervention cohort's written or oral exam was simpler than the control cohort in Dadaab.

As a result, we paid particular attention to specific questions in the Moodle quiz. Of 29 identical questions, 4 were significantly better for the control group and 2 for the intervention group. This result does not confirm our initial hypothesis that the intervention group scored better on average. However, on average, we had a double participation rate compared with the previous cohort, which can create a bias, as only the most motivated students from the control cohort participated in quizzes. Therefore, we looked further and considered the same questions in the written exam. There was a significant difference between the 2 scores in favor of the intervention group. In doing so, the inequity in the response rate is eliminated and ensures the same parameters. We interpret this improvement as emanating from the learning videos and interactive quizzes, making a difference to students' knowledge acquisition. Khan Academy videos give the effect of a live explanation, and quizzes with progressive levels help motivate students to learn—as testified by a participation rate close to 100%.

Improvement of Initial Knowledge Inside Groups

Determining whether the students had improved their initial medical knowledge during the course was difficult to assess. We inquired whether students had a background in the medical field before studying the course in the feedback form. Our analysis indicates that prior medical knowledge has a positive impact on the score, and a correlation was found between background knowledge and the final score in both groups. This could reveal 1 of 2 conditions: either students with previous health education learned something new or the group was too small to demonstrate a correlation. In either case, as R^2 adjusted was large, the size of the cohort does not change the correlation strength. Similarly, we compared the average of the quiz with the final result, which revealed a strong correlation. In the model, 87.3% of the final score variability was expressed as the average of the quiz. The excellent results of quizzes and final exams prove that their new knowledge was well received. We also established correlations among the exam questions.

Confounding Variables Regarding Final Scores

To identify any confounder that could enhance students' total scores, we first considered participation in the WhatsApp forum. We found no direct correlation between the number of WhatsApp messages posted by individual students and their final scores. However, the top scorer, S12, was also the most active on the WhatsApp forum, which came second for him as a study resource. In any case, in this challenging environment, we understand that not everyone has an electronic device and that the smartphones used to access WhatsApp were probably shared among students. However, we are reassured that the WhatsApp group is an excellent cohesive tool that gives all equal opportunities.

Second, we looked at the students' Moodle logs, which revealed no correlation. This reflects the basic activity of the Moodle platform but not engagement. With the new course organization focused on Khan Academy videos, the Moodle platform was relegated to be a tool where students could answer the quiz and evaluate their progress.

Finally, we noticed a significant negative correlation for gender, with women scoring 40% less than their male counterparts. Indeed, "traditional mechanisms of protection and social norms remain, but deviate substantially within the refugee context including the attitude and the perception of the 'proper' role of women. In some instances, this triggers a positive redefinition. In other cases, however, traditional mechanisms of protection and social norms remain in and this often results in stigmatization of the refugee woman by her community" [49].

With this in mind, we acknowledge that the lack of or reduced empowerment of female students negatively impacted their performance in the oral exam. This difference was not realized as strongly in the control group; the difference in score between males and females was 4.6% (P=.40) and not significant [33].

Impact

Of all participants, 10 students received accreditation from the University of Geneva (6 European Credit Transfer System [ECTS] credits) for module 1 and 1 student received a certificate of participation. Receiving ECTS credits allows students to transfer these credits into other medical training programs. The course has been discussed with local universities, and we aim to develop cooperation and future education pathways to advance students' studies.

One of the main problems identified with the Dadaab cohort project was the lack of critical thinking about the information they were learning. During the oral exam for the Kakuma cohort, we used a critical thinking question as a sensor to see if the students could apply what they learned to a new situation, in this case, a clinical problem. The results were presented as a bimodal distribution, which means that there is a clear separation of each student's ability to process what they learned. Some were perfectly capable of criticizing what they learned and using the information in a different context. Others needed guidance to do so, but they could be taught and learned. However, in a different educational system where knowledge is imparted, it is difficult to adjust the study approach.

This course also shows that there is flexibility in the learning system. All content is accessible via the Moodle platform and the tablets provided to the students at the start of the course. It is possible to take the course in other difficult refugee contexts. One of the students, for example, had to leave the camp and continued to attend the course and participated in the WhatsApp forum remotely. New resources and technologies could be incorporated into the course to further improve accessibility, which can be challenging in a refugee camp environment.

Finally, students told us orally and in their written feedback about their hopes to have a positive impact on their communities' health. They expressed their desire to improve the lives of the people around them by using their new knowledge. After a meeting with the International Rescue Committee (IRC), they confirmed that there is a need for more people to work in Kakuma's health care system. Consequently, the IRC expressed a desire to hire medical students in the camp's clinical facilities. In response, teachers have oriented the end of modules 2 and 3 to include some basic clinical gestures to help them achieve their objectives.

Limitation and Future Direction

The dropout rate in the newer cohort in Kakuma (5/16, 31%) was similar to that in the control group (9/27, 33%). The main reason for dropping out given by the students was *personal reasons*. In our view, although personal reasons cannot be anticipated at the time of recruitment, we recommend more clarity to the students before enrollment on the commitment needed to finish the course successfully.

The main criticisms for module 1 in both cohorts were the amount of time students needed to study. The multiplicity of content and tight deadlines were challenging for students. Even by making some schedule adjustments, the course is very demanding, and in the unstable context of a refugee camp, it can become difficult for some students to keep up with the required pace. Access to electricity, the internet, and more generally to the learning hub remains to be a problem for students in Kakuma refugee camps. The use of tablets may have been a way to ameliorate these challenges, with all the materials preloaded on the tablets. Unfortunately, for various security-related reasons, students could not take them home with them to study in the evenings and weekend days (eg, carrying valuable technology in the camp puts the students at risk of being attacked and robbed).

As with all web-based and blended learning courses, maintaining motivation is a challenge students face, which affects course completion rates. The education model used in Kakuma attempted to address this problem with a more interactive course to promote greater student involvement. The change in strategy to include students in the decision-making process and the heightened role of the onsite facilitator appears to have had a positive impact on the students' overall performance.

Future Directions

Going forward, we recommend evaluating the long-term impact of the program. A longitudinal study following the alumni to measure their impact on their communities and the opportunities they received would be a valuable resource for the growing field

of higher education in refugee contexts. Such a study would also help universities and other education providers operating in refugee spaces to refine their offers and better serve refugee students.

We also recommend that multiple courses are offered simultaneously and in several camps so that more refugees can attain the knowledge that they need to better ensure the health and well-being of their communities. Randomization would therefore be possible, thus ensuring greater internal validity. In addition, the introduction of a preliminary examination to assess the level of knowledge of course applicants more deeply would allow better organization of classes along capacity lines.

Finally, we suggest that any future courses for refugee learners consider the lived realities of refugee learners in refugee camps. Course content needs to reflect refugee needs and not be merely parachuted in via Western-centric digital modes. Blended learning will work if there are sufficient support and technological and pedagogical resources available for scaffold learning. To help achieve this, we recommend finding solutions so that students are safe and have greater access to technology. For example, giving them individual tablets at the start of the course and unlimited access to the learning hub could increase course participation and learning. In addition, contextualizing learning materials according to needs assessments and heightening the pedagogical and pastoral role of onsite facilitators will help to achieve better outcomes for all.

Conclusions

The InZone-Raft project has provided refugees in Kakuma and Dadaab refugee camps with high-level basic medical training in 2 of the most challenging environments in the world. We found that by improving the mode of delivery, better contextualizing content, and promoting more interaction between the students and their teachers, the Kakuma student cohort reached a high level of medical knowledge and was able to develop complex questions on medical topics. Their results in quizzes and final exams of their course prove that their new knowledge was well received, and the education model was more efficient than its earlier incarnation in the Dadaab refugee camp.

As we move forward in developing and delivering medical courses, those of us working with refugee learners need to pay more attention to our students. Given the exponential growth of the globally displaced in recent years, it is evident that the demand far outweighs the supply. However, this cannot be an excuse to cut corners and implement courses that do not work in refugee contexts. Much more positive results are achieved when refugees are acknowledged, included, and provided with the resources they need.

The positive outcome of this project shows that given the right resources and support, refugees have the capacity to contribute to, improve, and develop health care in their communities. However, to achieve this, a more concerted effort needs to be made to meet the educational needs of refugee learners and better understand the space in which they live. Higher education for refugee context research is one of the keys to achieve this. By better understanding the pedagogical dynamics that operate in these spaces, we can move forward with successful programs such as the InZone-Raft basic medical training course in Kakuma refugee camps.

The basic medical training course has been the subject of many requests from potential students and health care workers in Kakuma, as there is a lack of adequate education programs available in the camp. If there is the will to embrace it at the forced migration management administration level, this medical course could allow refugees to integrate into the health services in the camp. The positive impact this would have on the health and well-being of refugees in Kakuma could serve as inspiration that refugees can be empowered through higher education to serve and lead their own communities. Therefore, we recommend that similar education programs be embraced and implemented in refugee camps globally.

Conflicts of Interest

None declared.

Multimedia Appendix 1
InZone-Raft Basic Medical Training application form with an overview of the program, inclusion criteria, application form and motivational essay, and WhatsApp forum guidelines.
[PDF File (Adobe PDF File), 291 KB - jmir_v23i3e22345_app1.pdf]

Multimedia Appendix 2
The syllabus of Module 1 of the Basic Medical Training with unit number, names of organ systems, references of videos of the Khan Academy, references of book chapters, absolute numbers of questions in quizzes, average of quizzes per level, and duration of each unit in weeks.
[PDF File (Adobe PDF File), 92 KB - jmir_v23i3e22345_app2.pdf]

Multimedia Appendix 3
Complete characteristics of the intervention and control group participants including gender, prior knowledge, origins, and absolute frequency by group, mean, standard deviation, median, median absolute deviation, range, kurtosis, skew, and standard errors of the final scores broken down by gender and prior knowledge in both groups.

[PDF File (Adobe PDF File), 18 KB - jmir_v23i3e22345_app3.pdf]

Multimedia Appendix 4
WhatsApp group analysis of the module 1 intervention cohort was performed using WhatsAnalyzer. It includes general statistics, messages sent per user, messages sent on average during module 1, messages sent on average per day, messages sent on average per hour, and the communication network between users.
[PDF File (Adobe PDF File), 251 KB - jmir_v23i3e22345_app4.pdf]

Multimedia Appendix 5
Feelings were analyzed in the WhatsApp chat intervention by generating the global cloud, retrieving feeling words from the text, counting feeling words by category, printing the feeling plot, and creating a global text mining cloud. In addition, a graphical representation of the sentiment analysis is provided for 10 different sentiments in the WhatsApp intervention group.
[PDF File (Adobe PDF File), 145 KB - jmir_v23i3e22345_app5.pdf]

References

1. Pinheiro R, Wangenge-Ouma G, Balbachevsky E, Cai Y. The role of higher education in society and the changing institutionalized features in higher education. London: Palgrave Macmillan; 2015:225-242.
2. Papadimitriou A. Beyond rhetoric: reinventing the public mission of higher education. Tert Educ Manag 2019 Aug 3;26(1):1-4. [doi: 10.1007/s11233-019-09046-9]
3. Tilak JBG. International Handbook of Educational Research in the Asia-Pacific Region: Part One. In: Keeves JP, Watanabe R, Maclean R, Renshaw PD, Power CN, Baker R, et al, editors. Higher education and development. Dordrecht: Springer; 2003:809-826.
4. Higher education and the sustainable development goals. UNESCO. 2017. URL: https://en.unesco.org/themes/higher-education/sdgs [accessed 2020-11-24]
5. UNHCR Global Trends 2019. UNHCR - The UN Refugee Agency. URL: https://www.unhcr.org/statistics/unhcrstats/5ee200e37/unhcr-global-trends-2019.html [accessed 2020-11-24]
6. Amnesty International. Global facts and figures about refugees. URL: https://www.amnesty.org/en/what-we-do/refugees-asylum-seekers-and-migrants/global-refugee-crisis-statistics-and-facts/ [accessed 2020-11-24]
7. United Nations High Commissioner for Refugees. Kenya Statistics Package - 31 January 2019. URL: https://www.unhcr.org/ke/wp-content/uploads/sites/2/2019/02/Kenya-Statistics-Package-January-2019.pdf [accessed 2020-11-24]
8. USA for United Nations High Commissioner for Refugees. Protracted refugee situations explained. URL: https://www.unrefugees.org/news/protracted-refugee-situations-explained/ [accessed 2020-11-24]
9. Sustainable Development Goals Report 2020. United Nations Department for Economic and Social Affairs. 2020. URL: https://www.un.org/development/desa/publications/publication/sustainable-development-goals-report-2020 [accessed 2021-03-04]
10. Kirk J. Gender, forced migration and education: identities and experiences of refugee women teachers. Gender and Education 2010 Mar;22(2):161-176. [doi: 10.1080/09540251003606925]
11. Zeus B. Exploring barriers to higher education in protracted refugee situations: the case of burmese refugees in Thailand. J Refugee Stud 2011 Apr 28;24(2):256-276. [doi: 10.1093/jrs/fer011]
12. Taylor S, Sidhu RK. Supporting refugee students in schools: what constitutes inclusive education? Int J Inclusive Edu 2012 Jan;16(1):39-56. [doi: 10.1080/13603110903560085]
13. Wright LA, Plasterer R. Beyond basic education: exploring opportunities for higher learning in Kenyan refugee camps. Higher Education for Refugees 2012 Jan 18;27(2):42-56. [doi: 10.25071/1920-7336.34721]
14. Bellino MJ. Education, merit and mobility: opportunities and aspirations of refugee youth in Kenya's Kakuma refugee camp. Br Educ Res J 2020 Dec 13:-. [doi: 10.1002/berj.3695]
15. Molla T. African refugee youth in Australia: higher education participation. High Edu Res and Dev 2020 Dec 04:1-15. [doi: 10.1080/07294360.2020.1849035]
16. Akkari A, O'Keeffe P. University education in refugee camps must meet refugee needs. The Conversation. 2020 May 18. URL: http://theconversation.com/university-education-in-refugee-camps-must-meet-refugee-needs-137796 [accessed 2020-11-24]
17. Madianou M. Technocolonialism: digital innovation and data practices in the humanitarian response to refugee crises. Social Media + Society 2019 Jul 26;5(3). [doi: 10.1177/2056305119863146]
18. Halkic B, Arnold P. Refugees and online education: student perspectives on need and support in the context of (online) higher education. Learn, Media Tech 2019 Jul 16;44(3):345-364. [doi: 10.1080/17439884.2019.1640739]
19. Carron D. The study and teaching of human rights in refugee camps: learning from experience. URL: https://www.europenowjournal.org/2019/03/04/the-study-and-teaching-of-human-rights-in-refugee-camps-learning-from-experience/ [accessed 2020-11-24]

20. Dahya N. Education in conflict and crisis: how can technology make a difference? A landscape review. 2016. URL: https://www.eccnetwork.net/resources/education-conflict-and-crisis [accessed 2021-02-24]
21. Bawa P. Retention in online courses. SAGE Open 2016 Jan 05;6(1). [doi: 10.1177/2158244015621777]
22. O'Keeffe P. Flexible and adaptive responsiveness — Disruptive lessons from higher education in refugee contexts. 2020 Nov 22. URL: https://www.proudpen.com/book/the-impact-of-covid19-on-the-international-education-system-pp/flexible-and-adaptive-responsiveness-disruptive-lessons-from-higher-education-in-refugee-contexts/ [accessed 2020-11-24]
23. United Nations High Commissioner for Refugees. UNHCR-Kenya-Fact-Sheet-February-2019. 2019. URL: https://www.unhcr.org/ke/wp-content/uploads/sites/2/2019/04/UNHCR-Kenya-Fact-Sheet-February-2019.jpg [accessed 2019-04-11]
24. United Nations High Commissioner for Refugees. Kakuma refugee camp and Kalobeyei integrated settlement - UNHCR Kenya. UNHCR. 2019. URL: https://www.unhcr.org/ke/kakuma-refugee-camp [accessed 2019-04-11]
25. United Nations High Commissioner for Refugees. Investigation into allegations of refugee smuggling at the Nairobi Branch Office of the Office of the United Nations High Commissioner for Refugees: Note by the Secretary-General. UNHCR - The UN Refugee Agency. URL: https://www.unhcr.org/excom/unhcrannual/3deb346c5/investigation-allegations-refugee-smuggling-nairobi-branch-office-office.html [accessed 2019-04-11]
26. Licensed Health Facilities - Medical Practitioners and Dentists Board. URL: http://medicalboard.co.ke/online-services/licensed-health-facilities/ [accessed 2019-04-11]
27. Bolon I, Mason J, O'Keeffe P, Haeberli P, Adan HA, Karenzi JM, et al. One Health education in Kakuma refugee camp (Kenya): from a MOOC to projects on real world challenges. One Health 2020 Dec;10 [FREE Full text] [doi: 10.1016/j.onehlt.2020.100158] [Medline: 32844109]
28. Higher education for refugees - InZone - UNIGE. 2017. URL: https://www.unige.ch/inzone/what-we-do/history/higher-education-refugees/ [accessed 2019-04-16]
29. Stephane C. Pwr: basic functions for power analysis. R Package Version 1.3-0. 2020. URL: https://cran.r-project.org/web/packages/pwr/pwr.pdf [accessed 2021-02-24]
30. Thibault L. Course: Kakuma basic medical training 2018. 2020. URL: https://hop.unige.ch/course/view.php?id=17§ion=1 [accessed 2020-11-27]
31. Khan Academy: free online courses, lessons & practice. URL: https://www.khanacademy.org/ [accessed 2020-11-24]
32. OpenStax College. Anatomy and Physiology. Rice University. 2013. URL: https://openstaxcollege.org/files/textbook_version/hi_res_pdf/13/col11496-op.pdf [accessed 2019-04-08]
33. Burkardt AD, Krause N, Velarde RMC. Critical success factors for the implementation and adoption of e-learning for junior health care workers in Dadaab refugee camp Kenya. Hum Resour Health 2019 Dec 09;17(1):98 [FREE Full text] [doi: 10.1186/s12960-019-0435-8] [Medline: 31818294]
34. Revelle WR. Psych: procedures for personality psychological research. Northwestern University, Evanston, Illinois, USA. 2019. URL: https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-personality-and-psychological-research [accessed 2021-02-24]
35. RStudio Team. RStudio: integrated development for R. RStudio, PBC. Boston, MA. 2020. URL: https://www.Rstudio.Com/ [accessed 2021-02-24]
36. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2020:- [FREE Full text] [doi: 10.1201/b17972-15]
37. Signorell AEA. DescTools: tools for descriptive statistics. R Package Version 0.99.37. 2021. URL: https://cran.r-project.org/web/packages/DescTools/index.html [accessed 2021-02-24]
38. WhatsAnalyzer – App to analyse WhatsApp chats. URL: https://www.uni-wuerzburg.de/en/news-and-events/news/detail/news/whatsanalyzer-app-to-analyse-whatsapp-chats/ [accessed 2019-04-08]
39. Feinerer I, Hornik K. Tm: Text Mining Package. R Package Version 0.7-8. 2020. URL: https://cran.r-project.org/web/packages/tm/index.html [accessed 2021-02-24]
40. Fellows I. Wordcloud: word clouds. R Package Version 2.6. 2018. URL: https://cran.r-project.org/web/packages/wordcloud/wordcloud.pdf [accessed 2021-02-24]
41. Jockers ML. Revealing sentiment and plot arcs with the Syuzhet package. 2015. URL: https://www.matthewjockers.net/2015/02/02/syuzhet/ [accessed 2021-02-24]
42. Wickham H. Data analysis. In: ggplot2. Switzerland: Springer; 2016:189-201.
43. Microsoft Excel. Microsoft Corporation. 2018. URL: https://www.microsoft.com/en-in/microsoft-365/excel?legRedir=true&CorrelationId=0acd7117-b904-4ab4-ae65-af3f07258e76&rtc=1 [accessed 2021-02-24]
44. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch L. E1071: misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071). R Package Version 1.7-3. 2019. URL: https://cran.r-project.org/web/packages/e1071/index.html [accessed 2021-02-24]
45. Murdoch D, Chow ED. Ellipse: functions for drawing ellipses and ellipse-like confidence regions. R Package Version 0.4.2. 2020. URL: https://cran.r-project.org/web/packages/ellipse/index.html [accessed 2021-02-24]
46. Wuertz D, Setz T, Chalabi Y. FBasics: Rmetrics - markets and basic statistics. R Package Version 3042.89.1. 2020. URL: https://cran.r-project.org/web/packages/fBasics/fBasics.pdf [accessed 2021-02-24]

47. Jiahui W, Ruben Z, Alfio M, Victor Y, Matias S, Ricardo M, et al. Robust: port of the S+ robust library. R Package Version 0.5-0.0. 2020. URL: https://cran.r-project.org/web/packages/robust/robust.pdf [accessed 2021-02-24]
48. Martin M, Peter R, Christophe C, Valentin T, Andreas R, Matias S, et al. Robustbase: basic robust statistics. R Package Version 0.93-6. 2020. URL: https://cran.r-project.org/web/packages/robustbase/citation.html [accessed 2021-02-24]
49. United Nations High Commissioner for Refugees (UNHCR). Ensuring gender sensitivity in the context of refugee status determination and resettlement. Module 3: Ensuring Gender Sensitivity in the Context of Resettlement (Resource Package). 2005. URL: https://www.refworld.org/docid/43e73bef4.html [accessed 2019-04-09]

Abbreviations

ECTS: European Credit Transfer System
IRC: International Rescue Committee
MOOC: massive open online course
SDG: Sustainable Development Goal
UNHCR: United Nations High Commissioner for Refugees

Edited by R Kukafka; submitted 09.07.20; peer-reviewed by J Osero, M Manzanares; comments to author 17.11.20; revised version received 30.11.20; accepted 17.02.21; published 24.03.21.

Please cite as:
Lovey T, O'Keeffe P, Petignat I
Basic Medical Training for Refugees via Collaborative Blended Learning: Quasi-Experimental Design
J Med Internet Res 2021;23(3):e22345
URL: https://www.jmir.org/2021/3/e22345
doi:10.2196/22345
PMID:33759802

Original Paper

Linguistic Analysis of Online Communication About a Novel Persecutory Belief System (Gangstalking): Mixed Methods Study

Andrew Lustig[1,2*], MD, MSc; Gavin Brookes[3*], PhD; Daniel Hunt[4*], PhD

[1]Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[2]Centre for Addiction and Mental Health, Toronto, ON, Canada

[3]ESRC Centre for Corpus Approaches to Social Science, Department of Linguistics and English Language, Faculty of Arts and Social Sciences, Lancaster University, Lancaster, United Kingdom

[4]School of English Studies, Faculty of Arts, University of Nottingham, Nottingham, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Andrew Lustig, MD, MSc
Department of Psychiatry
Faculty of Medicine
University of Toronto
1051 Queen Street West
Toronto, ON, M6J1H3
Canada
Phone: 1 416 535 8501 ext 32841
Email: andrew.lustig@camh.ca

Abstract

Background: Gangstalking is a novel persecutory belief system whereby those affected believe they are being followed, stalked, and harassed by a large number of people, often numbering in the thousands. The harassment is experienced as an accretion of innumerable individually benign acts such as people clearing their throat, muttering under their breath, or giving dirty looks as they pass on the street. Individuals affected by this belief system congregate in online fora to seek support, share experiences, and interact with other like-minded individuals. Such people identify themselves as targeted individuals.

Objective: The objective of the study was to characterize the linguistic and rhetorical practices used by contributors to the gangstalking forum to construct, develop, and contest the gangstalking belief system.

Methods: This mixed methods study employed corpus linguistics, which involves using computational techniques to examine recurring linguistic patterns in large, digitized bodies of authentic language data. Discourse analysis is an approach to text analysis which focuses on the ways in which linguistic choices made by text creators contribute to particular functions and representations. We assembled a 225,000-word corpus of postings on a gangstalking support forum. We analyzed these data using keyword analysis, collocation analysis, and manual examination of concordances to identify discursive and rhetorical practices among self-identified targeted individuals.

Results: The gangstalking forum served as a site of discursive contest between 2 opposing worldviews. One is that gangstalking is a widespread, insidious, and centrally coordinated system of persecution employing community members, figures of authority, and state actors. This was the dominant discourse in the study corpus. The opposing view is a medicalized discourse supporting gangstalking as a form of mental disorder. Contributors used linguistic practices such as presupposition, nominalization, and the use of specialized jargon to construct gangstalking as real and external to the individual affected. Although contributors generally rejected the notion that they were affected by mental disorder, in some instances, they did label others in the forum as impacted/affected by mental illness if their accounts if their accounts were deemed to be too extreme or bizarre. Those affected demonstrated a concern with accumulating evidence to prove their position to incredulous others.

Conclusions: The study found that contributors to the study corpus accomplished a number of tasks. They used linguistic practices to co-construct an internally coherent and systematized persecutory belief system. They advanced a position that gangstalking is real and contested the medicalizing discourse that gangstalking is a form of mental disorder. They supported one another by sharing similar experiences and providing encouragement and advice. Finally, they commiserated over the challenges of proving the existence of gangstalking.

(J Med Internet Res 2021;23(3):e25722) doi:10.2196/25722

KEYWORDS

internet; discourse analysis; psychosis; delusions; linguistics; language; online discourse; corpus linguistics; computer mediated communication; schizophrenia; eHealth

Introduction

Gangstalking is a novel persecutory belief system whereby those affected believe they are being followed, stalked, and harassed by a large number of people, often numbering in the thousands [1,2]. In contrast to traditional forms of stalking that are usually organized by a single person [3], those affected by gangstalking are unable to identify a single person responsible for their persecution and experience it as a widely distributed and coordinated effort of co-conspirators. People who identify as affected by gangstalking self-identify as targeted individuals.

Although specific experiences of gangstalking vary between those affected, the various expressions of this polythetic belief system include a number of common elements. In particular, the campaign of harassment that affected individuals perceive is frequently experienced as an accretion of innumerable individually benign acts such as people clearing their throat, muttering under their breath, or giving dirty looks as they pass on the street. Perceived as deliberate, connected, and malicious, intense distress is experienced as a cumulative effect of these acts over a prolonged period. Individuals affected by gangstalking are frequently unable to pinpoint a clear motive for the harassment, which is a further source of perplexity and distress. They frequently describe that the apparent goal is to make them appear mentally ill, to cause them to be discredited and disbelieved, and sometimes to encourage or precipitate their eventual suicide.

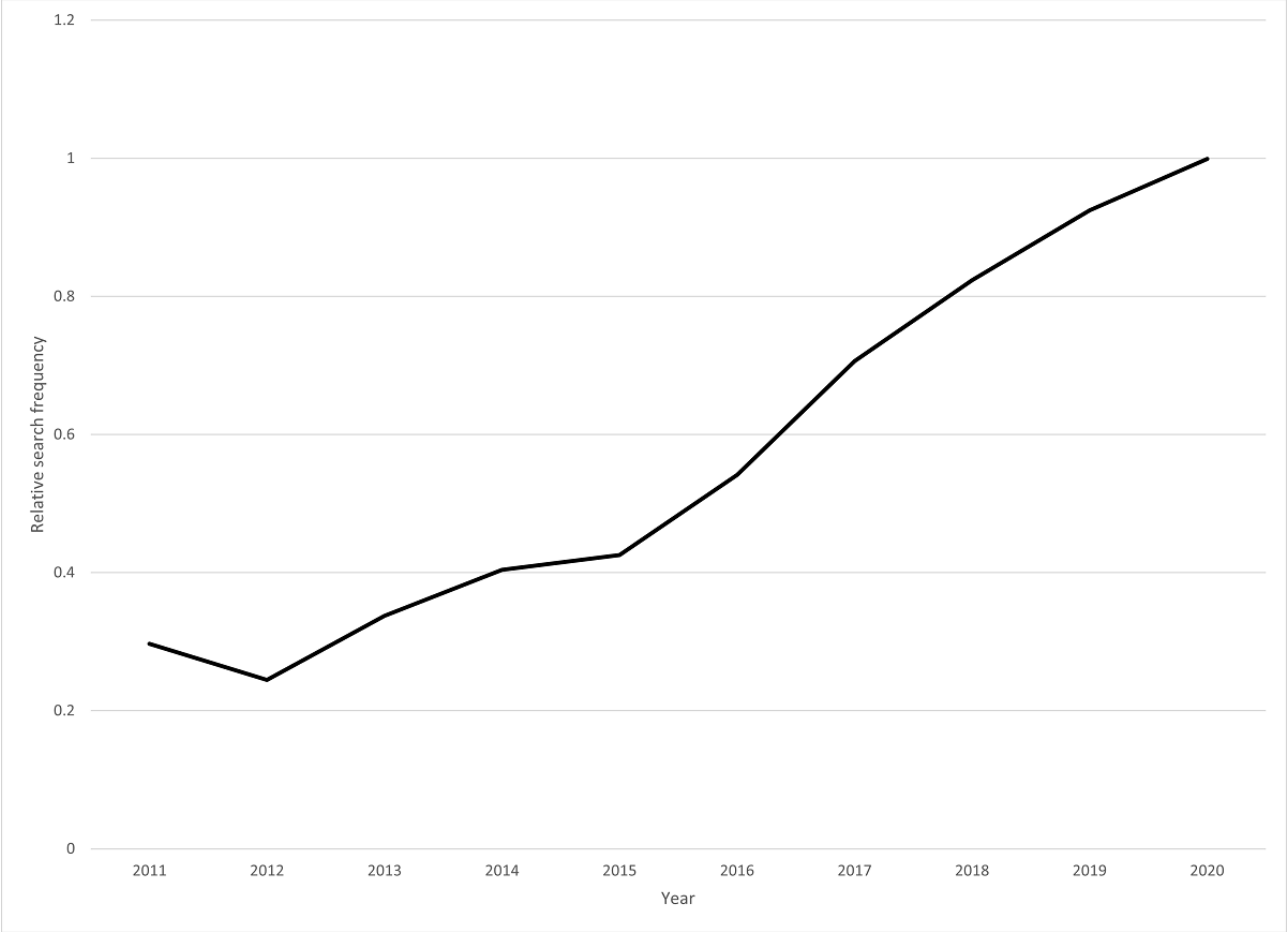
Interest in gangstalking is increasing over time and the popular press reports the activities of those affected with growing frequency [4-7]. As shown in Figure 1, the popularity of the Google search term *gangstalking* has increased steadily over the past decade [8]. When targeted individuals present to clinical attention, they are frequently diagnosed with psychotic illnesses and the gangstalking is conceptualized as a persecutory delusional system by psychiatric professionals. The gangstalking belief system is similar to some other well-established persecutory delusional belief systems, such as the *Truman Show* delusion [9], where those affected believe that their lives are surreptitiously being continuously recorded and produced into a reality television show and that everyone or nearly everyone they come into contact with is complicit in the deceit. As with many stigmatized beliefs [10,11], individuals affected by gangstalking reject the psychiatric formulation of their condition and turn elsewhere for support.

Targeted individuals congregate in online fora where they can speak openly of their concerns, flesh out their ideas, and comment on each other's experiences. These fora are a nonclinical environment where those affected may express their beliefs more openly and transparently without the fear of being disbelieved or labeled as may be the case in clinical settings. The internet has become an important source of health information [12]. In addition to providing a platform for those affected to find support, online fora may also serve as a crucible where people flesh out, develop, and linguistically and rhetorically construct the gangstalking phenomenon. It may also serve as a medium of transmission of the ideas as with other belief systems [13]. This study aims to describe how users of an internet forum about gangstalking construct, support, and contest the gangstalking belief system. It also seeks to describe how they use language to navigate social relationships within the context of the forum and as part of these processes.

Delusions are defined as fixed beliefs that are not amenable to change in light of conflicting evidence [14]. An alternative definition is that delusions are beliefs that are demonstrably untrue or not shared by others and which are not ordinarily accepted by other members of the person's culture or subculture [15]. However, attempts to precisely define delusions have proven problematic and debate and controversy persist [16], with some authors suggesting that pinning down delusions definitively may be an impossible task [17]. For example, superstitious beliefs resemble delusions and are widely held among people who are not affected by mental illness [18]. Other belief systems such as astrology, tarot, and parapsychology also resemble delusional belief systems, yet people endorsing these belief systems are not usually classified as experiencing delusions. Although there are widely accepted hypotheses regarding a biological underpinning of delusions, to date there is insufficient evidence to support a clear mechanistic explanation of them [19]. Moreover, the content of delusions varies across place and time and appears to be heavily influenced by prevalent cultural trends and symbols [20].

For these reasons, we regard persecutory belief systems and their variants such as conspiracy theories, overvalued ideas, and idiosyncratic belief systems as being socially constructed [21,22]. One of the key tenets of social constructionism is that knowledge is sustained by social processes [23]. This view holds that it is through discourse that certain, dominant ways of viewing and understanding particular phenomena come to be regarded as *truth*, at the expense of other perspectives [24]. It is on this basis (ie, through discourse) that certain psychological or embodied experiences come to be understood and treated within a society as being either *normal* or *pathological* and, by extension, those who experience that phenomenon as either *healthy*, *ill*, or even *deviant*. In this paper, we adopt a social constructionist approach to understanding the roles that language, discourse, and other social processes play in constructing the gangstalking phenomenon.

Figure 1. Relative frequency of "gangstalking" as a Google search term.



Methods

The methodology adopted in this study can be described as corpus-based discourse analysis. Corpus linguistics is largely a methodology (but also a field of research) which involves using computational techniques to examine recurring linguistic patterns in large, digitized bodies of authentic language data. Discourse analysis is an approach to text analysis which focuses on the ways in which linguistic choices made by text creators contribute to particular functions and representations. The approach to corpus-based discourse analysis employed in this study is derived from that described at length by Hunt and Brookes [25] in a previous analysis of mental health–related discourse in online fora.

This approach relies on a combination of 3 techniques from corpus linguistics: keyword analysis, collocation analysis, and manual examination of concordances. The first 2 techniques are quantitative methods that use statistical techniques to sift through a large body of text (known as a corpus) to identify, respectively, words and word combinations that are notable due to their high frequency or statistical salience [26]. The third technique, concordancing, is essentially a way of viewing the corpus data that allows users to inspect all instances of a given word, word string, or collocational pairing in the corpus—in context—and, if it is desired, to access the original corpus texts in their entirety. Concordancing facilitates more qualitative analysis of the patterns in a corpus. In this study, it is used to follow up the identification of keywords and collocational pairing, with the ensuing qualitative analysis trained on identifying the wider discursive and rhetorical practices that the keywords and collocates signal and through which the forum users construe their relationships, identity, and experiences.

To obtain source texts for our corpus of forum interactions about gangstalking, we used Google to identify support groups for people experiencing gangstalking. We then focused on the largest gangstalking forum on the internet in terms of number of users, threads, and posts. The forum used to construct the study corpus is organized into topics, each one of which has an accompanying discussion which forms a thread. We used Python 3.0 code to extract 420 complete threads (225,936 words; see Table 1). The data collected included all threads posted between July 17, 2020, and September 2, 2020 (the date of collection). Some threads that were posted and subsequently deleted by their authors were not available for analysis. This was the case for 80 of the 500 threads we sought to extract, which left 420 threads for analysis. The forum requires posters to successfully solve a CAPTCHA before posting to prove they are human and not a bot.

All of the data used in our analysis were posted on a public forum, available to any internet user without having to subscribe or log into the forum. The forum permits users to contribute anonymously with a pseudonymous username that is not linked to their offline identities. Our examination of the forum posts constitutes what Eysenbach and Till [27] refer to as passive

analysis. The institutional research ethics board at The Centre for Addiction and Mental Health reviewed the proposed study design and opined that it did not require formal approval.

To help preserve contributors' anonymity, we term this corpus the *gangstalking internet corpus*. At the time of data collection, the gangstalking forum that we sampled had a total of approximately 14,000 (exactly 13,598) members. To ensure that forum members' identities are protected as far as is possible, no usernames or references to any other personally identifying information will be reproduced in the data extracts cited in this paper.

Our analysis began by using version 8 of the corpus analytical software *WordSmith Tools* [28] to identify keywords in the study corpus. Keywords are words that occur in the study corpus with a statistically marked frequency when compared with a reference corpus, which usually represents a norm or benchmark for the type of language under study [29]. As our reference corpus, we elected to use the spoken component of the updated British National Corpus [30]—an 11-million-word corpus of conversational British English sampled between 2012 and 2016. This reference corpus was also used by Hunt and Brookes [25], who demonstrated its utility for identifying keywords which signal discursive and rhetorical practices in the context of online fora.

Keyness was measured using a combination of the log-likelihood and the log ratio statistic [31]. Log-likelihood is a confidence measure. The higher the log-likelihood value assigned to a keyword, the smaller the probability that the (statistically marked) observed frequency of that word has arisen due to chance or a sampling error, for example. Log ratio, by contrast, is an effect size measure. The higher the log ratio score assigned to a particular keyword, the larger the observed difference is between its frequencies in the analysis corpus and the reference corpus. We stipulated that keywords should have a log-likelihood score of 15.13, indicating a confidence level of 99.99%. We also specified that a word had to be present in at least 2.5% of forum posts (ie, 69 posts out of 2749) in order to be identified as a keyword. We then ranked the resulting keywords using the log ratio statistic [32]. We set a minimum log ratio of 1.5 for a word to be included as a keyword. A log ratio of 1.5 means that the word is 2.25 times as frequent in the study corpus as in the reference corpus.

After identifying keywords, we grouped them into thematic and semantic categories. We began with the categories defined by Sheridan et al [2] in their content analysis of self-defined gangstalking-affected individuals' accounts of their subjective experiences of the phenomenon and modified them to capture the themes that emerged from our keyword list.

Following keyword categorization, we extracted collocates of a select number of keywords of interest, in order to examine the wider linguistic contexts within which those words tended to occur in the forum posts. This step takes us beyond the solitary items in the keyword output and begins to move toward understanding the meanings and functions of words in context; as Firth [33] puts it, "you shall know a word by the company it keeps." In this way, collocate analysis can identify the meanings and associations that affected individuals attribute to different aspects of gangstalking. We defined collocates as words occurring within 5 words to the left or right of the search word (this is the default in *WordSmith Tools* and had been found to be productive for corpus-based discourse studies, eg, by Hunt and Brookes [25]; Tables 2 and 3). Collocation was measured and ranked using the cubed version of the mutual information (MI) statistic (MI^3). The MI^3 statistic highlights collocational pairings whose frequency is marked (ie, higher than would be expected given the frequencies of the constituent words and the size of the corpus overall). It is useful for corpus-based discourse analysis, as it favors high-frequency collocational pairings which are thereby particularly well established in the discourse [34]. For analyses of computer-mediated communication, this offers the practical advantage that it does not place undue emphasis on infrequent collocates that are typos or spelling errors.

Finally, keywords and collocational pairings of interest were subjected to manual discourse analysis using concordance output and, where beneficial, based on entire forum posts and those which precede and follow them in the threads. As noted, the objective of this stage of the analysis was to identify the discursive and rhetorical practices through which the forum contributors construed the gangstalking phenomenon and their experiences of it.

Table 1. Profile of the gangstalking internet corpus.

Attribute	Gangstalking corpus
Total threads sampled	420
Total posts	2749
Mean posts per thread	6.54
Total words	225,836
Mean words per post	82.1

Table 2. Top 5 collocates of gangstalking (5 left/5 right), ranked by MI3[a].

Rank	Collocate	Frequency	MI^3
1	*the*	107	16.59
2	*and*	66	14.74
3	*that*	41	13.70
4	*you*	31	12.38
5	*are*	25	12.09

[a]MI^3: cubed version of the mutual information (MI) statistic.

Table 3. Top 5 lexical collocates of gangstalking, ranked my MI3[a].

Rank	Collocate	Frequency	MI^3
1	*people*	14	10.32
2	*believe*	12	11.53
3	*think*	12	10.77
4	*One*	11	10.37
5	*Real*	11	12.04

[a]MI^3: cubed version of the mutual information (MI) statistic.

Results

Study Analysis Overview

As described in the previous section, we began our analysis by obtaining keywords from our corpus of gangstalking forum threads. We modified Sheridan and James [1] initial 24 thematic categories of the gangstalking experience to 9 aggregate keyword categories (Table 4): (1) conceptions of gangstalking, (2) social and interpersonal concepts, (3) conceptualizations of the individual, (4) mental and psychological processes, (5) epistemic indicators, (6) extent of conspiracy, (7) technological affordances employed in gangstalking, (8) words pertaining to the internet, and finally, (9) grammatical words were categorized together. Some words were assigned to multiple categories. For example, the polysemous word *state* can refer to a state of mind. It can also refer to a nation or political community. For this reason, it was placed in 2 categories.

Table 4. Keyword categories.

Thematic/lexical category	Associated keywords ranked by log ratio score (frequencies [n] in brackets)
Conceptions of gangstalking	*Gangstalking* (380), *gangstalkers* (148), *perps* (113), *gangstalked* (83), *stalkers* (152), *stalked* (99), *stalking* (347), *targeted* (183), *target* (146), *program* (95), *TI*[a] (144), *evil* (91), *control* (195), *situation* (88)
Social and interpersonal concepts	*Harassment* (161), *victim* (137), *gang* (278), *torture* (141), *other* (152), *power* (111), *involved* (130), *against* (181), *social* (101), *anyone* (225), *help* (223), *group* (127)
Conceptualizations of the individual	*Victim* (137), *individuals* (98), *individual* (104), *human* (110), *life* (345), *person* (275), *someone* (298), *myself* (115)
Mental and psychological processes	*Fear* (120), *mental* (128), *state* (102), *believe* (352), *experience* (121), *crazy* (110), *mind* (276)
Epistemic indicators	*Evidence* (128), *information* (129), *believe* (352), *happening* (94), *real* (190), *reason* (126), *seem* (83), *seems* (100)
Extent of conspiracy	*Government* (200), *public* (120), *police* (176), *state* (102), *law* (115), *using* (130), *world* (240)
Technological affordances	*Technology* (131), *video* (91)
Internet related	*https* (411), *www* (242), *com* (271), *lol* (72), *post* (121)
Grammatical	*etc (167)*, *its* (235), *themselves* (114), *become* (90), *may* (188), *am* (335), *being* (600), *by* (795), *their* (851), *also* (415), *most* (293), *without* (131), *will* (741), *since* (122), *case* (92)

[a]TI: targeted individuals.

Lexicalizing Gangstalking

The keywords belonging to the category *Conceptions of gangstalking* illustrate that those affected employ various lexical choices for constructing gangstalking in their forum posts. Comparing raw frequencies, it is most commonly referred to as *gangstalking*, which occurs 380 times in the corpus, and *stalking*, which occurs 347 times. *Gangstalking* is the gerund form of *gangstalk*, a portmanteau of *gang* and *stalk*. The word is a neologism. It is not included in standard English language dictionaries and indeed is absent from the updated Spoken British National Corpus, which served as our reference corpus for the keyword analysis above. Gangstalking is sometimes lexicalized as the bigram *gang stalking* in our corpus (n=142 occurrences). The words *stalking* (n=347) and *harassment* (n=161) were also used.

The term *gangstalking* served several different functions in our corpus. In some instances, it serves as a progressive verb. In other instances it is used as a gerund or as a present participle and functions as an adjective. For example:

> *LIL WAYNE IS **GANGSTALKING** AND HARASSING ME*
>
> *The **gangstalking** scumbags at the bottom of the hierarchy are usually exploited and disrespected endlessly.*
>
> *But they still play the childish **gangstalking** games.*

While the first of these examples demonstrates that gangstalking is conceived of as a process similar to harassment (and, in this case, perpetrated by a famous musician), examples 2 and 3 demonstrate the way in which the existence of gangstalking is frequently represented as presupposed and uncontroversial. That is, the use of *gangstalking* as a descriptor of people or games functions as an existential presupposition; the use of gangstalking in this way presupposes it is. This implies that gangstalking is a valid and real concept.

The determiner *the* is the most frequent collocate of *gangstalking* in our corpus, occurring in the L1 position (ie, immediately to the left of *gangstalking*) a total of 47 times in the corpus. Lexicalizing *gangstalking* with the definite article *the* frames it entirely as an entity external to the affected individual. Moreover, use of the definite article indicates the verbal, as opposed to the nominal, gerund which conceptualizes a specific and actualized situation that is marked as identifiable [35]. The forum serves as a site of discursive contest between 2 competing worldviews. According to one, the concerns about gangstalking reside within the affected individual as part of a medicalizing discourse. In this paradigm, the experience of gangstalking may be regarded as a chemical imbalance or psychological disturbance. The countervailing view, by contrast, adopts a credulous persecutory discourse and posits that the difficulty is entirely due to the thoughts and behavior of malevolent others located outside of the affected individual. Use of the determiner *the* anticipates this contest and supports the latter view, which is a minority discourse in psychiatric practice, but the majority in this corpus.

Below are several examples of this construction.

> *Satan is definitely at work when it comes to **the gangstalking** and he is using technology as well as gang stalking perps as human vessels to get his will accomplished.*
>
> *Even I filed complaints to Federal, provincial, and other organizations, **the gangstalking** increases.*
>
> *The **gangstalking** was heavy. Every day. Every minute of the day. I still didn't know what it was. I thought it was bullying, and I ""deserved"" it for being different.*

These comments speak about affected individuals' concern that the phenomenon is widespread, insidious, and centrally coordinated. Much like the use of *gangstalking* as a presupposition, these examples also demonstrate how gangstalking is represented as taking place regardless of the affected individual's perceptions. This is achieved through the linguistic process of nominalization, in which the process of gangstalking is presented as a noun (the gangstalking). In the second extract above (ie, *Even I filed complaints …*), for instance, it is not that the affected individual *perceives* that the gangstalking is becoming more intense or that they are being gangstalked more frequently, rather their post expresses the seemingly objective fact that their gangstalking has *increased*, once again presenting the phenomenon as incontrovertible.

Representations of the gangstalking phenomenon invariably include references to the perpetrators of the conspiracy as well as affected individuals. In the gangstalking community, the victims are usually known as *targeted individuals*. TI occurred 144 times in the data, *targeted* 183 times, and *individuals* 98 times, while *targeted individual* and *targeted individuals* occurred a total of 86 times. Among those affected, the perpetrators of gangstalking are known as *perps*. That word occurred 113 times in the corpus.

The next most frequent collocate of *gangstalking* is the coordinating conjunction *and*, which occurs most commonly in the R1 position (ie, directly following the node).

> *I have **gangstalking and** direct energy weapons/remote neural monitoring happening to me.*
>
> *Can confirm Iran not exempt from **gangstalking and** very advanced mind control technologies.*
>
> *This **gangstalking, and** chronic chemical poisonings, have taken a toll on my health.*
>
> *Undoubtedly, these Government stalking worthless punks could not afford nice cars, lavish homes, and domestic fees, if they were not **gangstalking and** research people's brains 24/7.*

As these examples attest, forum contributors use the conjunction to situate the gangstalking behavior within a matrix of similar persecutory and malicious behaviors. In this manner, members of this community construct gangstalking as an individual phenomenon that is intertwined with broader national and international conspiracies. This includes 14 references to "direct energy weapons" and 69 references to voice to skull (V2K) communication technologies. As an online phenomenon, this may also increase contact between individuals who experience

gangstalking as an aspect of a persecutory delusion and members of internet conspiracy cultures more generally.

Consistent with this claim, the most frequent keyword in our analysis was *https*, appearing 411 times across 140 of the comments. It was used in the context of URLs, pointing readers to other resources on the internet that commenters used to emphasize and elaborate on their ideas. This speaks to the hyperlinked and connected nature of the internet and online communities but also to the nature of gangstalking as a belief system that has been popularized and shared through the networked communication of the web.

Several of the keywords highlighted interpersonal themes. This demonstrates that the gangstalking belief system is based on malicious interpersonal interactions. Affected individuals identify themselves as *victim* (n=137) and use the term *gang* (n=278) to describe their tormentors. The words *someone* (n=298) and *anyone* (n=225) are used to describe people involved in the conspiracy. Both are indefinite pronouns and allow for doubt about who they are describing. This may speak to the inchoate nature of the belief system in which those affected are certain that they are being targeted even if they cannot always precisely pinpoint whom by.

Figures of Authority

One of the keyword categories pertains to the broad reach of the conspiracy. These words identify powerful state actors. Almost without exception, those affected cannot identify a single person or agent who is responsible for their persecution. Some affected individuals construe the gangstalking as being retribution for a minor slight or altercation in the past. However, they speak of agents of symbolic authority such as *government* (n=200), *police* (n=176), and *state* (n=102) as either having an organizing role or at least permitting and encouraging the persecution.

> *Gangstalking definitely is coming from **government**, but it is using the private sector to avoid detection.*
>
> *The **government** has been using this technology to target specific people and also experimentally torture some people.*
>
> *Since **police** is involved in this, there is very little we can do about it and sueing them won't help but don't let that deter you.*
>
> *I have a history of being stalked by the **police** so, no, I don't ask them for help.*

This formulation frames gangstalking as a process that is occurring outside the affected individual. However, throughout the corpus, contributors also refer to the alternative view that gangstalking may be a psychological process. The keywords *believe*, *mental*, *mind*, and *experience* all draw attention to the epistemological and ontological challenges faced by those affected: what is really happening and how can one be sure? For example:

> *Do you actually **believe** in this?*
>
> *They will not **believe** you. It would be crazy to **believe** us without evidence.*
>
> *Also, describing gangstalking will often sound completely illogical - people will not **believe** the government would spend that much time or money on a person.*

Individuals affected by gangstalking express concern about being able to demonstrate the veracity of their experiences. *Evidence* occurs 128 times in the corpus. It most frequently occurs with the collocates *collect* (MI^3=15.41) and *gather* (MI^3=14.18). Those affected post about the need, the challenges, and the potential benefits of accumulating sufficient evidence to conclusively demonstrate the veracity of the belief system:

> *If you are not presenting some form of **evidence** to skeptics, you are wasting your time.*
>
> *It sounds like you have the opportunity to gather **evidence** and confirmation this is happening.*

Faced with the risk of being disbelieved, being portrayed as mentally ill is a central concern of those affected. A common theme running through their accounts is that the very purpose of the campaign is to discredit and stigmatize them by making them appear chronically mentally ill. For example, the adjective *crazy* occurs 110 times in the corpus with almost all instances pertaining to their concerns about being labeled as mentally ill and stigmatized:

> *Yeah it's they ritual to drive you **crazy** so you act weird so they can put you as **crazy** person so nobody will listen to the abuse.*
>
> *The trick is to be subtle so you don't come across as **crazy** or threatening.*
>
> *I was told at the beginning they would make me look **crazy** or lying so no one would believe it.*

Throughout the corpus, individuals impacted by gangstalking deal with the possibility and the assertion that they are affected by mental illness. Throughout the corpus, those impacted deal with this tension and the possibility that they are affected by mental illness by representing *craziness* as the intended outcome of gangstalking that they are actively resisting. Accordingly, those affected rarely acknowledge that they have mental illness. However, in some cases posters posit that other forum members do.

> *You guys are actually insane...*
>
> *Hey man you need serious mental help.*

In other instances, posts note that it is actually the perpetrators of gangstalking that are affected by mental illness:

> *Most Government gangstalkers dispatched to you, have severe psychological problems, and are afflicted with a serious mental illnesses.*
>
> *It is not your fault for being gang stalked. Since stalkers have mental illness or personality disorder that fuels this behaviour.*

In this manner, references to mental illness in the community serve to insulate the majority of its members from the contention that they themselves are affected by delusions. Mental illness is seen as a characteristic of *perps* rather than targeted

individuals or is attributed to a small number of community members whose experiences are dismissed as too extreme.

The frequent use of the word *seem* (n=83) and its variations, *seems* (n=100), *seemed* (n=24), and *seemingly* (n=13), could be viewed as reflecting uncertainty relating to aspects of affected individuals' accounts of their gangstalking experiences. However, tellingly, these linguistic markers of uncertainty did not reflect any uncertainty relating to the legitimacy of gangstalking itself. Rather, the forum members used *seem* and its related forms to hypothesize about the nature of their own or others' gangstalking experiences, as well as to theorize about its effects on them as individuals. As the next example demonstrates particularly well, such hypothetical scenarios tend to err on the side of the *gangstalking* explanation for the experiences being described, thereby arguably bolstering the legitimacy of the phenomenon.

> *...what **seemed** to be the same man, although I couldn't get a good look at him.*
>
> *It **seems** like once I feel a great level of peace, they come around to bring me down.*
>
> ***Seems** like you are being gangstalked by an actual gang.*

Discussion

It is well established that online social support confers mental health benefits upon patients [36,37]. However, the contested nature of gangstalking makes the role of this forum more ambiguous. On the one hand, the forum offers a platform for those affected by gangstalking to be heard and believed, in some instances without the stigma of being labeled as mentally ill. On the other hand, in some instances the forum may serve to further reinforce a maladaptive belief system, drawing those affected further into an echo chamber or down the rabbit hole of conspiracy, reinforcing previously held beliefs and discouraging them from seeking treatment.

Our analysis identified a lexicon comprising words that are highly salient to members of the gangstalking community, many of which are likely to be unfamiliar to outsiders. This includes words like *gangstalking* itself, as well as words that label the various actors in the gangstalking universe including *targeted individuals* and *perp*. In addition, contributors use specialized vocabulary to describe technological affordances such as V2K to describe "voice to skull" technologies to broadcast sounds into the minds of those affected. In addition to its communicative function, using these words serves to validate and legitimize forum contributors as members of the community [38].

The data depict the forum as a site of ontological discursive contest between 2 opposing worldviews about the nature of gangstalking. In one, it is seen as a widespread, crowdsourced system of persecution involving many members of the community, the government, police, and other figures of symbolic authority. The countervailing view is that it is a product of mental disorder and a figment of affected individuals' imaginations. The linguistic practices in this corpus show that *gangstalking* is lexicalized in various ways that take its existence as given. In addition to constructing and representing coordinated harassment as an objective state of affairs, the nominalization of *gangstalking* also obscures the agent of the harassment and the party affected by it. Contributors use *seem* and its variants to hedge and capture a sense of uncertainty. Further, though gangstalking includes a core set of beliefs [1,2], individual expressions of the belief system vary from person to person. The term *gangstalking* allows forum contributors with varying experiences to have a common nomenclature to refer to their experiences for the purposes of exchanging stories and support with alike others. Moreover, it might be argued that the label *gangstalking* provides the forum users with a means with which to confer a sense of symbolic order over a set of otherwise incoherent experiences, in the process perhaps granting them a sense of control over it [39], or at the very least the linguistic apparatus with which to convey their distress and seek out others who are "in the same boat."

Despite the potential value of labeling and naming a contested phenomenon like gangstalking, it is nevertheless important to note that many of the contributors to this also manifested a concern about being labeled as mentally ill and generally rejected such a formulation. Although the contributors acknowledge that the distress caused by persecution, alienation, and disbelief may be a source of psychological distress and mental disorder, they also reject the formulation that the belief system is itself a product of the mind. However, some descriptions of gangstalking that are deemed too extreme are labeled as pathological by other group members, which implies that these members operate with a vaguely specified gradient along which experiences of gangstalking may be classed as being pathological at one end and not at all pathological on the other.

Our analysis highlighted the interpersonal nature of the belief system and affected individuals' concern with interpersonal processes. The gangstalking belief system is characterized by malice perpetrated by a vast number of unnamed others. These include private citizens and also official bodies such as police and government.

These results have the potential to inform clinicians interacting with patients who experience persecutory belief systems. Building a therapeutic relationship to enable engagement is the central process in therapy for psychosis [40]. Having a detailed understanding of the belief systems held by people affected by persecutory belief systems may be important in developing empathy and building a therapeutic alliance. Cognitive behavioral approaches to the treatment of persecutory belief systems recommend that clinicians partner with patients to critically evaluate and dispute delusional and other unhelpful beliefs. Doing so requires a detailed understanding of the beliefs and evidence for and against them. Our hope is that this study may be helpful in that regard.

Our study focused on a particular persecutory belief system. However, insights from this work may be applied more broadly to other, related belief systems. This analysis is particularly valuable because it is based on discussions taking place in a nonclinical setting which arguably allows for more candid and authentic communication, alleviating a potential "Hawthorne effect" of data collected in clinical settings.

Online fora such as the one examined here represent popular avenues for health-related support and advice seeking. This is likely the case, to an extent, for all health-related issues. Yet, this is particularly relevant to contested health issues such as gangstalking, whose contested clinical status may result in those affected turning to peers rather than practitioners for advice and social support. For practitioners seeking to learn about the belief systems and (patient) community norms associated with contested health issues, it therefore behooves them to become acquainted with such online peer support contexts and the linguistic routines (and associated discourses) that characterize the interactions that take place within them.

Acknowledgments

The following individuals merit thanks for providing helpful comments: David Chartash, David Gratzer, Jason Joannou, Erene Stergiopoulos, Albert Wong, and Juveria Zaheer.

Conflicts of Interest

None declared.

References

1. Sheridan LP, James DV. Complaints of group-stalking (‘gang-stalking’): an exploratory study of their nature and impact on complainants. The Journal of Forensic Psychiatry & Psychology 2015 Jun 16;26(5):601-623. [doi: 10.1080/14789949.2015.1054857]
2. Sheridan L, James D, Roth J. The Phenomenology of Group Stalking ('Gang-Stalking'): A Content Analysis of Subjective Experiences. Int J Environ Res Public Health 2020 Apr 06;17(7):2506 [FREE Full text] [doi: 10.3390/ijerph17072506] [Medline: 32268595]
3. Spitzberg BH, Cupach WR. The state of the art of stalking: Taking stock of the emerging literature. Aggression and Violent Behavior 2007 Jan;12(1):64-86. [doi: 10.1016/j.avb.2006.05.001]
4. Mcphate M. United States of Paranoia: They See Gangs of Stalkers. The New York Times. 2016 Jun 10. URL: https://www.nytimes.com/2016/06/11/health/gang-stalking-targeted-individuals.html [accessed 2020-10-21]
5. Weinberger S. Mind Games New on the Internet: a community of people who believe the government is beaming voices into their minds. They may be crazy, but the Pentagon has pursued a weapon that can do just that. The Washington Post. 2007 Jan 14. URL: https://tinyurl.com/s7ndk9uk [accessed 2020-10-21]
6. Pierre J. Gang stalking: Real-life harassment or textbook paranoia? Psychology Today. 2020 Oct 20. URL: https://www.psychologytoday.com/blog/psych-unseen/202010/gang-stalking-real-life-harassment-or-textbook-paranoia [accessed 2020-11-21]
7. Tait A. "Am I going crazy or am I being stalked?" Inside the disturbing online world of gangstalking. MIT Technology Review. 2020 Aug 7. URL: https://www.technologyreview.com/2020/08/07/1006109/inside-gangstalking-disturbing-online-world/ [accessed 2020-11-11]
8. Google Trends. URL: https://www.google.com/trends [accessed 2021-02-16]
9. Gold J, Gold I. The "Truman Show" delusion: psychosis in the global village. Cogn Neuropsychiatry 2012 Nov;17(6):455-472. [doi: 10.1080/13546805.2012.666113] [Medline: 22640240]
10. Fox N, Ward K, O'Rourke A. Pro-anorexia, weight-loss drugs and the internet: an "anti-recovery" explanatory model of anorexia. Sociol Health Illn 2005 Nov;27(7):944-971 [FREE Full text] [doi: 10.1111/j.1467-9566.2005.00465.x] [Medline: 16313524]
11. Baker D, Fortune S. Understanding self-harm and suicide websites: a qualitative interview study of young adult website users. Crisis 2008 May;29(3):118-122. [doi: 10.1027/0227-5910.29.3.118] [Medline: 18714907]
12. Cohen RA, Adams PF. Use of the Internet for health information: United States, 2009. In: NCHS Data Brief No. 66. Hyattsville, MD: National Center for Health Statistics; 2011.
13. Lustig A, Mackay S, Strauss J. Morgellons disease as internet meme. Psychosomatics 2009;50(1):90. [doi: 10.1176/appi.psy.50.1.90] [Medline: 19213978]
14. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5). Philadelphia, PA: American Psychiatric Association; 2013:22.
15. World Health Organization. International classification of diseases for mortality and morbidity statistics (11th Revision). 2018. URL: https://icd.who.int/browse11/l-m/en [accessed 2021-02-15]
16. Oyebode F. Sims' Symptoms in the Mind: An Introduction to Descriptive Psychopathology. Edinburgh, UK: Elsevier; 2008.
17. David A. On the impossibility of defining delusions. Philosophy, Psychiatry, & Psychology 1999;6(1):17-20.
18. Levitt EE. Superstitions: Twenty-Five Years Ago and Today. The American Journal of Psychology 1952 Jul;65(3):443. [doi: 10.2307/1418765]

19. Sadock B, Sadock V. Kaplan and Sadock's Synopsis of Psychiatry: Behavioral Sciences/Clinical Psychiatry (10th edition). Philadelphia, PA: Lippincott Williams & Wilkins; 2012.
20. Kala A, Wig N. Delusion across cultures. Int J Soc Psychiatry 1982 Jul 27;28(3):185-193. [doi: 10.1177/002076408202800304] [Medline: 7118457]
21. Georgaca E. Reality and discourse: a critical analysis of the category of 'delusions'. Br J Med Psychol 2000 Jun;73 (Pt 2):227-242. [doi: 10.1348/000711200160444] [Medline: 10874481]
22. Harper DJ. Delusions and Discourse: Moving Beyond the Constraints of the Modernist Paradigm. Philosophy, Psychiatry, & Psychology 2004;11(1):55-64. [doi: 10.1353/ppp.2004.0041]
23. Burr V. Social Constructionism (3rd Edition). New York, NY: Routledge; 2015.
24. Burr V, Butt T. Psychological distress and post-modern thought. In: Pathology and the Postmodern: Mental Illness as Discourse and Experience. London, UK: Sage; 1999:186-206.
25. Hunt D, Brookes G. Corpus, Discourse and Mental Health. London, UK: Bloomsbury Publishing; May 28, 2000.
26. McEnery AM, Wilson A. Corpus Linguistics: An Introduction. Edinburgh, UK: Edinburgh University Press; 2001.
27. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001 Nov 10;323(7321):1103-1105 [FREE Full text] [doi: 10.1136/bmj.323.7321.1103] [Medline: 11701577]
28. Scott M. WordSmith Tools Version 8. Stroud: Lexical Analysis Software; 2020.
29. Seale C, Ziebland S, Charteris-Black J. Gender, cancer experience and internet use: a comparative keyword analysis of interviews and online cancer support groups. Soc Sci Med 2006 May;62(10):2577-2590. [doi: 10.1016/j.socscimed.2005.11.016] [Medline: 16361029]
30. Love R, Dembry C, Hardie A, Brezina V, McEnery T. The Spoken BNC2014. IJCL 2017 Dec 1;22(3):319-344. [doi: 10.1075/ijcl.22.3.02lov]
31. Lüdeling A, Ungeheuer G, Steger H, Wiegand HE, Burkhardt A. Corpus Linguistics: An International Handbook. Berlin, Germany: De Gruyter; 2009.
32. Hardie A. ESRC Centre for Corpus Approaches to Social Science (CASS). 2014 Apr 28. URL: http://cass.lancs.ac.uk/log-ratio-an-informal-introduction [accessed 2021-02-15]
33. Firth J. Papers in Linguistics, 1934-1951. Oxford, UK: Oxford University Press; 1957.
34. Brezina V, McEnery T, Wattam S. Collocations in context. IJCL 2015 Aug 17;20(2):139-173. [doi: 10.1075/ijcl.20.2.01bre]
35. Maekelberghe C. The English gerund revisited. Corpus Linguistics and Linguistic Theory 2019 Apr 27;15(1):205-237. [doi: 10.1515/cllt-2016-0054]
36. Ziebland S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? Milbank Q 2012 Jun;90(2):219-249 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00662.x] [Medline: 22709387]
37. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
38. Wenger E. Communities of Practice: Learning, Meaning, and Identity. Cambridge, UK: Cambridge University Press; 1998.
39. Brookes G, Harvey K, Chadborn N, Dening T. “Our biggest killer”: multimodal discourse representations of dementia in the British press. Social Semiotics 2017 Jun 30;28(3):371-395. [doi: 10.1080/10350330.2017.1345111]
40. Kingdon D, Turkington D. Cognitive Behavioral Therapy of Schizophrenia. Hove: Psychology Press; 2000.

Edited by R Kukafka; submitted 13.11.20; peer-reviewed by JW Tan; comments to author 04.12.20; revised version received 11.12.20; accepted 08.02.21; published 05.03.21.

Please cite as:
Lustig A, Brookes G, Hunt D
Linguistic Analysis of Online Communication About a Novel Persecutory Belief System (Gangstalking): Mixed Methods Study
J Med Internet Res 2021;23(3):e25722
URL: https://www.jmir.org/2021/3/e25722
doi:10.2196/25722
PMID:33666560

Original Paper

Toward a Multivariate Prediction Model of Pharmacological Treatment for Women With Gestational Diabetes Mellitus: Algorithm Development and Validation

Carmelo Velardo[1,2], PhD; David Clifton[1,2], DPhil; Steven Hamblin[1], PhD; Rabia Khan[1], PhD; Lionel Tarassenko[1,2], DPhil; Lucy Mackillop[1,3,4], MA

[1]Sensyne Health, plc, Oxford, United Kingdom

[2]Department of Engineering Science, University of Oxford, Oxford, United Kingdom

[3]Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom

[4]Nuffield Department of Women's Reproductive Health, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Carmelo Velardo, PhD
Sensyne Health, plc
Schrodinger Building
Heatley Road
Oxford, OX4 4GE
United Kingdom
Phone: 44 0330 058 1845
Email: carmelo.velardo@sensynehealth.com

Abstract

Background: Successful management of gestational diabetes mellitus (GDM) reduces the risk of morbidity in women and newborns. A woman's blood glucose readings and risk factors are used by clinical staff to make decisions regarding the initiation of pharmacological treatment in women with GDM. Mobile health (mHealth) solutions allow the real-time follow-up of women with GDM and allow timely treatment and management. Machine learning offers the opportunity to quickly analyze large quantities of data to automatically flag women at risk of requiring pharmacological treatment.

Objective: The aim of this study is to assess whether data collected through an mHealth system can be analyzed to automatically evaluate the switch to pharmacological treatment from diet-based management of GDM.

Methods: We collected data from 3029 patients to design a machine learning model that can identify when a woman with GDM needs to switch to medications (insulin or metformin) by analyzing the data related to blood glucose and other risk factors.

Results: Through the analysis of 411,785 blood glucose readings, we designed a machine learning model that can predict the timing of initiation of pharmacological treatment. After 100 experimental repetitions, we obtained an average area under the receiver operating characteristic curve of 0.80 (SD 0.02) and an algorithm that allows the flexibility of setting the operating point rather than relying on a static heuristic method, which is currently used in clinical practice.

Conclusions: Using real-time data collected via an mHealth system may further improve the timeliness of the intervention and potentially improve patient care. Further real-time clinical testing will enable the validation of our algorithm using real-world data.

(J Med Internet Res 2021;23(3):e21435) doi:10.2196/21435

KEYWORDS

gestational diabetes mellitus; mobile health; machine learning; algorithms

Introduction

Background

Gestational diabetes mellitus (GDM), defined as glucose intolerance with onset or first recognition during pregnancy [1], increases the risk of morbidity in women and newborns. Successful management of GDM reduces maternal hyperglycemia and perinatal morbidity. Having a pregnancy complicated by GDM is associated with a risk of developing type 2 diabetes in the long term for women [2] and, later in life,

for the offspring [3]. Standard clinical management for GDM is lifestyle advice and pharmacological treatment [4,5].

The increased prevalence of GDM [6], combined with limited resources available to the National Health Service (NHS) [7], is challenging the optimal delivery of care to women with GDM in the NHS.

GDM is usually diagnosed after 24 weeks of pregnancy, providing only a short intervention period (typically around 10 weeks) to influence perinatal outcomes. Therefore, timely identification of the need for pharmacological treatment is very important.

In current clinical practice, blood glucose (BG) data are provided by women in the form of paper-based diaries that are bought to clinics for clinician review. Decisions regarding GDM management therefore occur only during these reviews.

Benefits of Mobile Health

Access to real-time data recorded in electronic diaries could enable *between-clinic* recognition of patterns in the data and allow midwives to adjust medication in a timely fashion so that women have more chance of tighter control of their BG readings, facilitating improved clinical outcomes for the woman and her baby.

Mobile health (mHealth; ie, internet-linked mobile devices to monitor patients' health) is well placed to facilitate this type of care and provide health care providers access to a larger and richer set of data on which to base their clinical decisions [8].

The use of mHealth by women with GDM can simplify their daily routine and has been shown to provide an easy and acceptable way of collecting, storing, and analyzing their data to aid personal decision making [9,10].

Digitally monitored BG can also provide additional benefits to clinicians by enabling real-time reviews and customized feedback. Furthermore, using the collected data could lead to the development of algorithms for the early identification of the need for pharmacological treatment, allowing earlier intervention, more frequent reviews, and potentially improved outcomes.

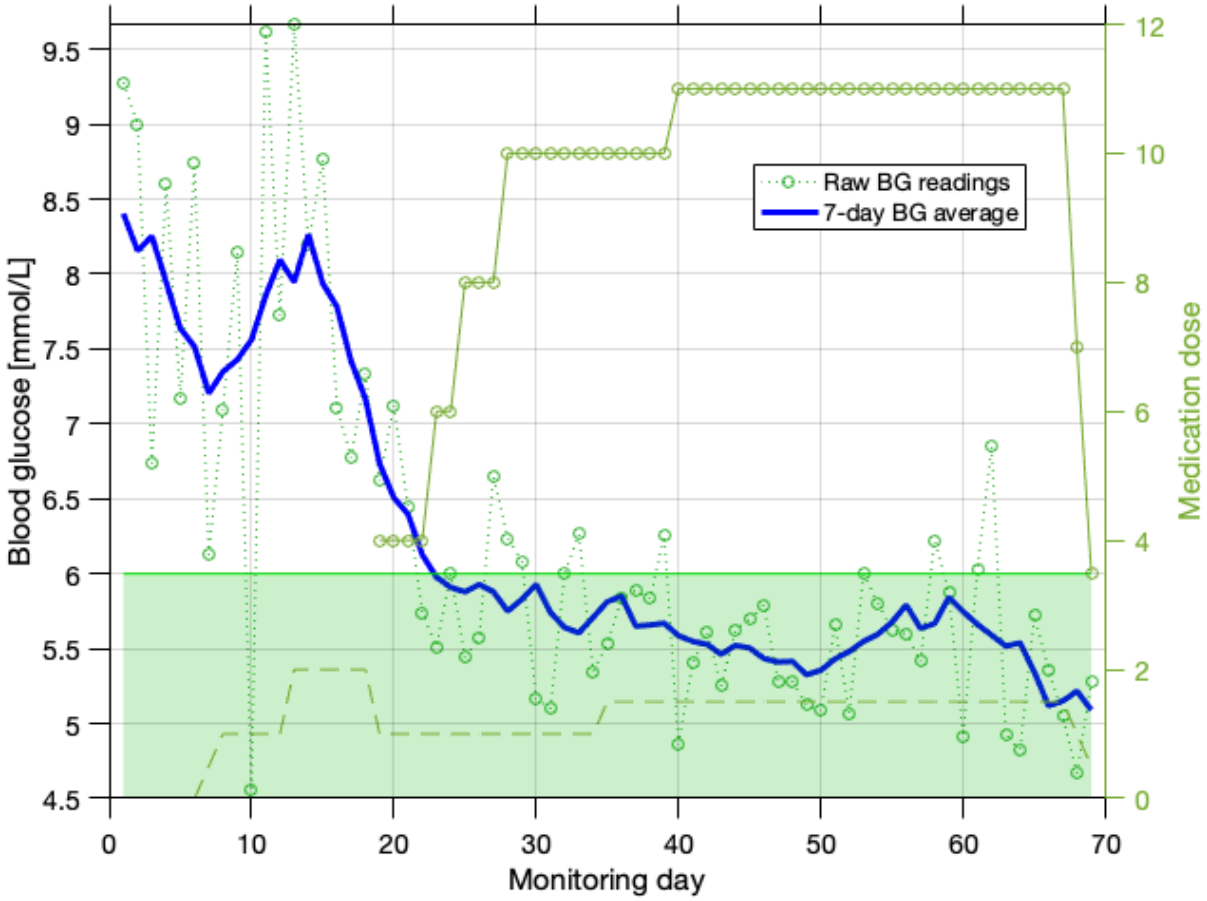
Some studies have successfully used baseline characteristics to predict the need for pharmacological treatment or maternal outcomes. Among these studies, factors such as glucose tolerance test results, maternal age, maternal BMI, ethnicity, and previous pregnancy history were identified as predictors of the requirement for pharmacological therapy [11-16].

Prediction of Pharmacological Treatment

Barnes et al [12] analyzed a cohort of 3317 women with GDM to predict pharmacological treatment using variables from electronic patient records. The authors validated their model using data from different time periods (eg, 1992-2004 and 2005-2015) and multiple clinics (Bankstown-Lidcombe Hospital Diabetes Centre and Liverpool Hospital Diabetes Centre). Their algorithm was capable of a positive predictive value of 87.6%, negative predictive value of 69.9%, sensitivity of 93%, specificity of 99.4%, and the area under the receiver operating characteristic curve (AUC) value of 0.712 (95% CI 0.693-0.731). This study makes use of variables at diagnosis, such as the oral glucose tolerance test and glycated hemoglobin (HbA_{1c}) results, to predict the need for pharmacological treatment. In contrast to our study, the authors do not include operational, real-time BG readings, which may be fundamental to obtain better predictions.

Data collected by mHealth platforms (an example is given in Figure 1) will enable the design of algorithms for the identification of women at risk (stratification) and the early detection of required pharmacological treatment (prediction). By using machine learning (ML) to analyze data, it will be possible to provide real-time feedback to women and clinicians and allow the development of decision-supported processes for the titration of medication therapy for hyperglycemia.

Figure 1. Patient data sample illustrating blood glucose readings, 7-day average blood glucose (BG), medication information, and target range (green shading). Both the dashed and solid circled lines represent medications. In this instance, the use of appropriate medication and dosage allows BG to reach the target range. BG: blood glucose.



This paper aims to describe the application of ML techniques to real-world data collected with an mHealth app to predict future medication events. This study compares how the predictive algorithm performs against standard heuristic techniques currently employed in NHS trusts and provides initial insights on future lines of work that will improve the existing model.

Methods

Data Collection

The GDm-Health system (Sensyne Health, plc) was used in research and clinical practice settings, and at the time of writing, it is currently routinely used for clinical practice in 35 NHS trusts.

The GDm-Health system was used to track (1) BG readings and (2) adherence to medications prescribed by health care professionals. Using the GDm-Health app, participants entered their readings, tagged them with information identifying the meal (eg, prebreakfast, postlunch, etc) and recorded information concerning the dose of any medication taken. Whenever enabled, wireless transfer from Bluetooth-enabled BG monitors was used; as an alternative (eg, if there were issues with the wireless transfer or use of a noncompatible meter), manual input of BG readings was also employed.

The analysis was performed on fully anonymized data based on established partnerships with these trusts.

For the analysis in this paper, one source of data was an implementation study that included data from (1) John Radcliffe Hospital, Oxford University Hospitals (OUH) NHS Foundation Trust, and (2) Royal Berkshire Hospital, Royal Berkshire Hospitals (RBH) NHS Foundation Trust.

This implementation study was performed by the Institute of Biomedical Engineering at the University of Oxford and OUH.

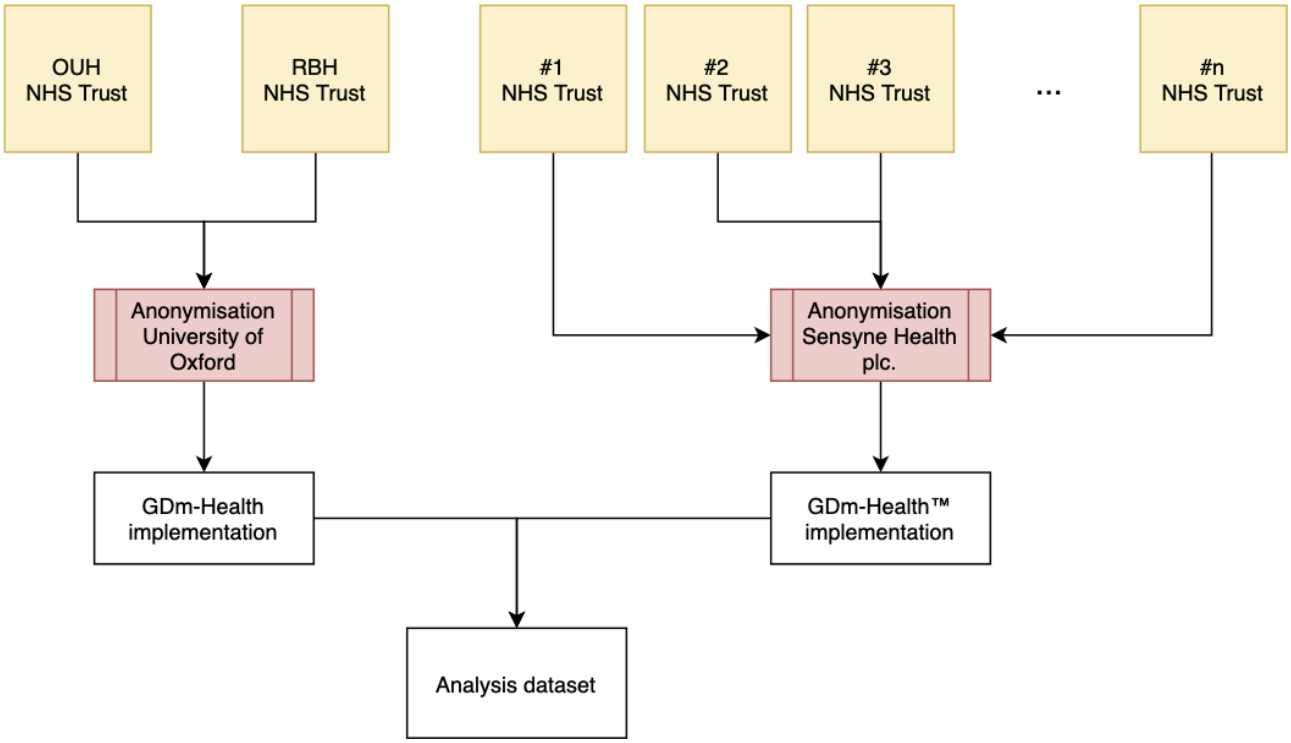
Data from the research implementation were collected for the period January 2016 to January 2019 for OUH and September 2014 to September 2019 for RBH.

A second larger set of data was generated by the anonymized data set collected and curated by Sensyne Health, plc, via the GDm-Health system. Anonymization was performed according to guidelines [17] and using a publisher (Privitar, London, United Kingdom).

Both data sets were analyzed in anonymized form. Anonymization included the removal of identifiable information (eg, given name, family name, NHS number) and transformation of sensitive information (eg, date of birth was transformed into age in years). In both instances (research and production implementation), the systems were used for clinical management

of women with GDM; therefore, this analysis corresponds to a retrospective, secondary care–based cohort study of women with GDM using the GDm-Health system (Figure 2 describes the data flow that contributed to the data set used for the analysis).

Figure 2. Data collection and anonymization flow. NHS: National Health Service; OUH: Oxford University Hospitals; RBH: Royal Berkshire Hospitals.



Pregnant women with GDM used the GDm-Health app to track pre- and postprandial BG. All women with GDM during the monitoring period were included in the analysis. Women with type 1 and type 2 diabetes and Maturity onset diabetes of the young (MODY) were excluded from the analysis.

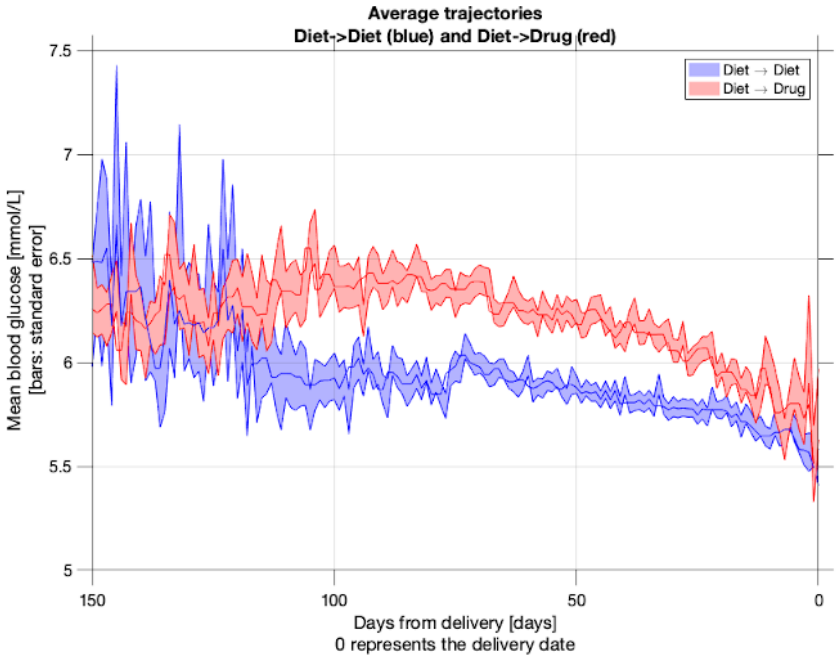
GDM was diagnosed using a variety of methods, including the International Association of Diabetes and Pregnancy Study Groups criteria [18] and the National Institute for Health and Care Excellence 2010 guidelines [19]. Clinical management included hospital-based follow-up in antenatal clinics and remote monitoring of BG readings using the GDm-Health app. Monitoring and management of patients varied across sites and included management via diet alone, metformin, insulin, or a combination of the above.

Analysis

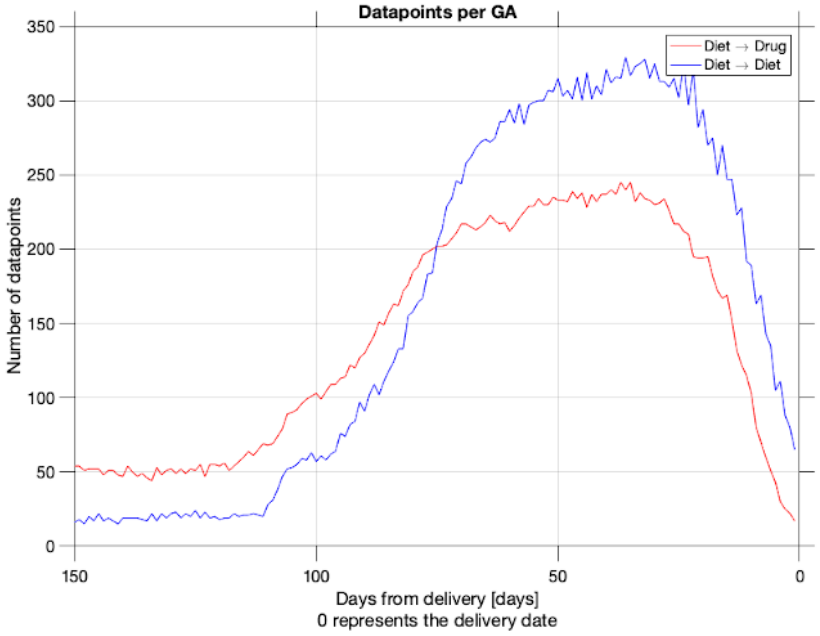
To develop an algorithm that could predict women in need of pharmacological treatment, we identified 2 subgroups of interest: the diet–diet group and the diet–drug group. Women belonging to the first group did not have any prescribed medication; therefore, they remained on lifestyle-based therapy throughout their pregnancy. Those belonging to the second group transitioned from lifestyle-based therapy to pharmacological treatment (metformin and/or insulin). We identified patients as belonging to this group when not taking any medication for at least 14 days from the first day of use of the GDm-Health system. If participants initiated pharmacological treatment before the cutoff period (2 weeks), they were excluded from the analysis as they would have initiated pharmacological treatment too early.

By aligning data from all involved women to their respective delivery date allows us to plot the average BG values regardless of their meal tag and their standard error per day up to delivery. Figure 3 shows the plot of the average BG per day up to the delivery date, and at the same time it shows the number of readings used to compute the statistics of the previous. One can observe how both averages decrease toward the delivery date, suggesting that treatment is successful in both groups. However, the average BG value for those in the diet–drug group was higher than that in the diet–diet group. This result points to the importance of the BG value in deriving features to distinguish women belonging to these 2 groups.

Figure 3. Graph (a) shows the average and SE of blood glucose values per day corresponding to the 2 groups. Graph (b) indicates how many readings per day were used to obtain the average in (a). In both cases, the number of days on the x-axis refers to the number of days to delivery. For the period where the number of readings is high, 100 to 5 days to delivery, which roughly corresponds to the last 3 months, the average of the blood glucose values for the 2 groups is clearly different.



(a)



(b)

Clinical Variables

To train a model that is capable of recognizing women belonging to one of the 2 specified groups (diet–diet and diet–drug), we trained an ML algorithm over a set of features (predictors) to be extracted from the training periods associated with the 2 groups.

A list of relevant predictors, informed by clinicians, was drawn up to summarize the monitoring period extracted from the data of each group. Wherever possible, this set of predictors was extended by considering the average and SD of variables over the monitoring period to capture the level and variability of each.

Table 1 lists the set of predictors used for our analysis, together with a detailed explanation of their nature, including the time and the way in which they were recorded. The majority of predictors describe the BG level at different times of the day (eg, prebreakfast, postdinner); 2 identify a consecutive alerting situation (eg, 3 days with high readings at the meal tag in a row), and 2 describe demographics of the patient (age and BMI).

Table 1. Description of predictors used in this study.

Feature	How it was expanded	When it was recorded	How it was recorded
Breakfast readings	Mean, SD, min[a], max[b], linear regression coefficient	Recordings were made according to the GDM[c] management plan each participant discussed with their health care professional	BG[d] data were recorded by participants through the GDm-Health mobile app developed to help mothers-to-be to keep track and manage their BG throughout their pregnancy
Lunch readings	Mean, SD, min, max, linear regression coefficient	Recordings were made according to the GDM management plan each participant discussed with their health care professional	BG data were recorded by participants through the GDm-Health mobile app developed to help mothers-to-be to keep track and manage their BG throughout their pregnancy
Evening meal readings	Mean, SD, min, max, linear regression coefficient	Recordings were made according to the GDM management plan each participant discussed with their health care professional	BG data were recorded by participants through the GDm-Health mobile app developed to help mothers-to-be to keep track and manage their BG throughout their pregnancy
All readings (regardless of prandial tag)	Mean, SD, min, max, linear regression coefficient	Recordings were made according to the GDM management plan each participant discussed with their health care professional	BG data were recorded by participants through the GDm-Health mobile app developed to help mothers-to-be to keep track and manage their BG throughout their pregnancy
Raised 3-day prebreakfast	Variable indicating the number of prebreakfast alerts in a 3-day consecutive period	Recordings were made according to the GDM management plan each participant discussed with their health care professional	BG data were recorded by participants through the GDm-Health mobile app developed to help mothers-to-be to keep track and manage their BG throughout their pregnancy
Raised 3-day postprandial	Variable indicating the number of alerts in a 3-day consecutive period for each postprandial meal	Recordings were made according to the GDM management plan each participant discussed with their health care professional	BG data were recorded by participants through the GDm-Health mobile app developed to help mothers-to-be to keep track and manage their BG throughout their pregnancy
BMI	BMI at the time of booking	Data were recorded at the first encounter with the health care professional	BMI is a derived variable from weight and height. Both these variables were recorded at the time of booking
Age	Age at the time of booking	Data were recorded at the first encounter with the health care professional	Variable derived, for privacy reasons, from the date of birth recorded at booking. Recorded as age in number of months
Ethnicity risk factor	Asian, Chinese, Pakistani, Bangladeshi, and other ethnicities considered at risk for GDM	Data were recorded at the first encounter with the health care professional	Data were manually recorded by the health care professional
Risk factors	Previous GDM, high BMI, family history of diabetes, previous large-for-gestational-age baby, ethnicity, polyhydramnios, glycosuria, macrosomia, missed OGTT[e], unable to take OGTT, polycystic ovaries	Data were recorded at the first encounter with the health care professional	Data were manually recorded by the health care professional

[a]min: minimum.

[b]max: maximum.

[c]GDM: gestational diabetes mellitus.

[d]BG: blood glucose.

[e]OGTT: oral glucose tolerance test.

Data collected were first analyzed at the population level. BG data distributions were inspected, and sensible (data-driven) thresholds were established to filter out readings that considered outliers. BG monitoring is highly affected by how the test is

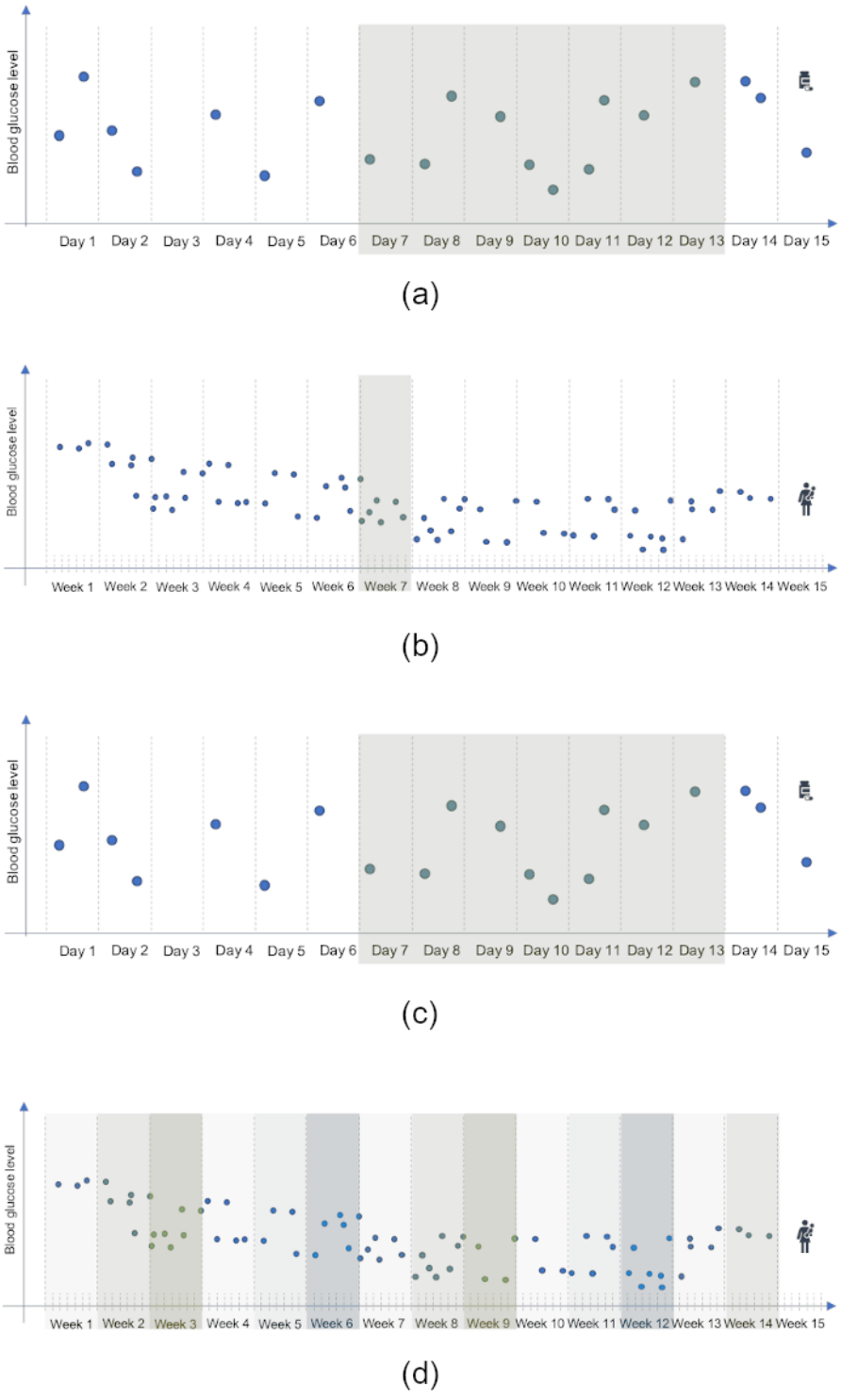
performed and the experience level of the user. Therefore, inconclusive tests can lead to skewed BG values that are not representative of the real BG levels, and such outliers must be identified and removed to avoid bias in the training data. For the BG level, the 95th percentile of all population data was selected as the cutoff threshold, and values above this level were excluded from the data set as outliers.

As we cannot assume that our analysis would be unaffected by missing data (eg, we cannot guarantee that data are missing entirely at random), variables with missing data were imputed by substituting each missing feature with the values of the population mean for that characteristic.

Development of an ML Model

Figure 4 shows the learning tasks considered. BG metrics and medication information were available on the GDm-Health platform. Medications are prescribed by health care professionals and are collected and available on the GDm-Health platform as self-reported information by women. Both the type of medication and its dosage were captured in the platform; however, only a binary representation (medication/no medication) was used during the analysis. For the diet–drug group, the training period considered corresponded to the week before the first medication was administered. For the diet–diet group, as it was challenging to identify a clear event and to train over a range of data that represent the whole predelivery monitoring period, a randomly selected week was chosen from all the available ones. For the diet–drug group, we excluded from the training week one day before the start of medication, whereas there was no need for a gap day for the diet–diet group.

Figure 4. For both (a and b) training and (c and d) validation, summary features are computed on 1 week's data and a label is assigned according to the group the data belong to (diet–drug and diet–diet). During training, (a) for women with a medication prescription (diet–drug), the week before the first medication dosage is considered for training. A 1-day gap between the training week and the medication event is maintained. (b) For women with no medication prescription (diet–diet), a random week is selected among all those available. During validation, (c) for women in the diet–drug group, we used a similar approach to training and computed the features on the week before the first day of medication (leaving a 1-day gap before the medication event). (d) For women in the diet–diet group, instead, we considered each week available for testing as an independent sample on which to perform a prediction.



Validation was performed similarly to the training for the diet–drug group, that is, based on features computed from the week before the medication event. For the diet–diet group, all available nonoverlapping weeks were independently considered. For each of these, a set of features was generated with the appropriate label for the diet–diet group.

To develop the statistical model, we trained a logistic regression model on training data and tested the output of the training model on the held-out validation data.

We repeated this experiment on 100 different random permutations of the main data set between training and validation data using a 70% training and 30% validation split. At each iteration, to avoid biasing the algorithm toward the overrepresented class (diet–diet), this was randomly downsampled to the number of women in the underrepresented class (diet–drug). The validation set was not downsampled, thereby retaining within it the class imbalance that would be observable in real data. Before training, a lasso feature selection step was performed on the training data and predictors deemed important to this step were selected and used for training the algorithm. The lasso MATLAB (MathWorks) function was used with its *alpha* parameter set to .75 (corresponding to elastic net regression) and using 5-fold cross-validation.

At each iteration, a different set of women would compose the training and testing set, thus training is performed on subjects who do not contribute data to the testing set.

The results of the test were evaluated by computing the receiver operating characteristic curve (ROC), which plots the false-positive rate (FPR) against the true-positive rate (TPR). The AUC was also computed to permit the comparison of different models and to provide a summary of the performance of the algorithm. From the 100 repeated experiments, a summary description of the ROC and AUC was obtained by calculating percentiles at 5%, 50%, and 95%, providing the median ROC curve and CIs at 5% and 95%, respectively.

To compute the ROC curves, risk groups were defined automatically by the perfcurve function (MATLAB 2019a) by varying the value of the decision threshold over the range of values from 0 to 1 produced by the logistic regressor. Finally, comparison with the standard of care was evaluated by visualizing the performance of the current methodology against the ROC curve. The current clinical heuristic states that treatment should be considered if 3 or more consecutive BG readings of the same meal tag are over the designated threshold.

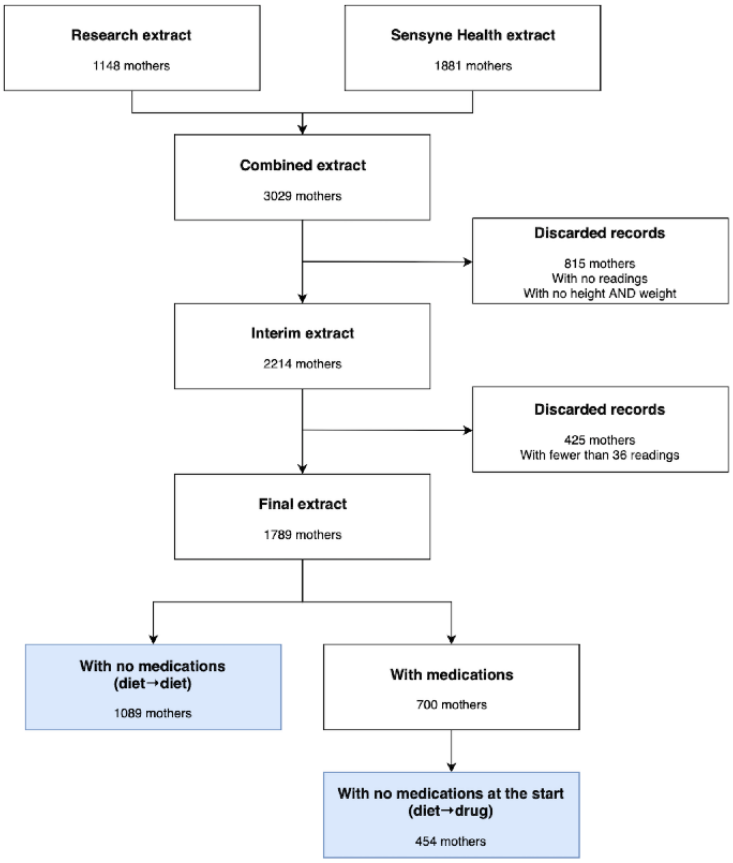
Results

Data Description

Data from 12 trusts were collected from women being monitored using the GDm-Health platform during 2019 (Sensyne Health, plc data set) and from the Oxford University research data set between 2014 and 2015.

A total of 3029 women were included in our data set at the time of analysis. After excluding data from women with incomplete demographic information (eg, missing weight and height) and excluding data from women with insufficient BG data (eg, women with fewer than 36 BG readings in the first week of use), data from 1789 women were analyzed. After further reduction and classification into one of the 2 groups of interest for our analysis (diet–diet and diet–drug), the remaining group of women (Figure 5) provided 411,785 BG readings (mean 230, SD 181), of which 160,812 were tagged as breakfast readings, 117,887 as lunch readings, and 133,086 as evening meal readings.

Figure 5. Consort diagram describing the data selection flow. Final groups are identified by the colored boxes.

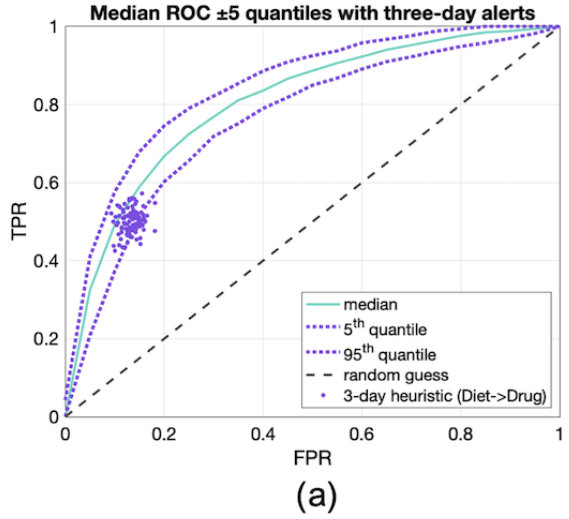


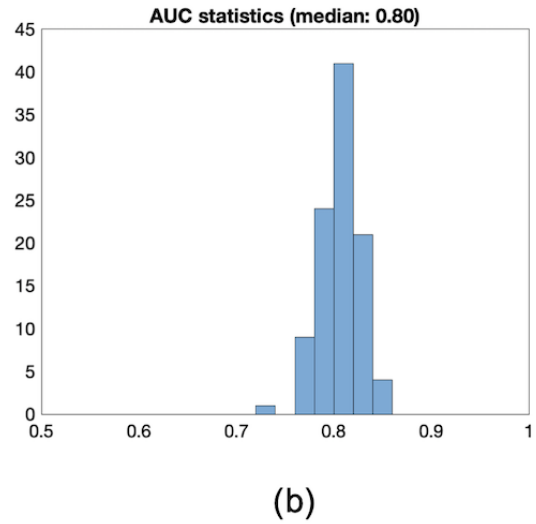
Among the 1789 patients, 39.12% (700/1789) women required pharmacological treatment during their pregnancy, whereas the remaining group was managed only through diet adjustments.

Analysis

The results of the analysis outlined in the Methods section are summarized in Figure 6.

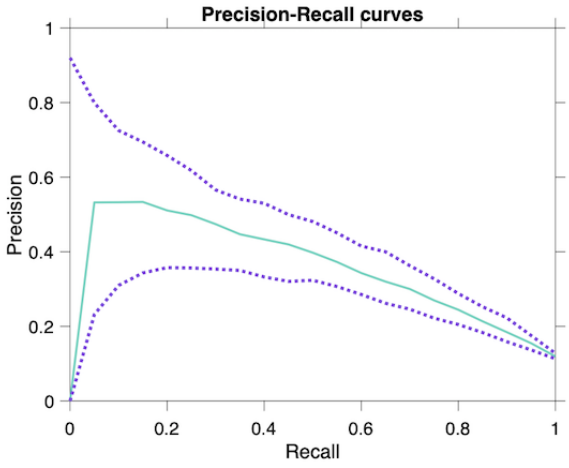
Figure 6. Experimental results. (a) ROC curves depicting the 5th and 95th centiles out of 100 repetitions of the classification experiment. (b) Histogram showing the distribution of area under the curve values for all 100 experiments. (c) Top 10 variables out of 100 repeated experiments. (d) Precision-recall curve depicting the 5th and 95th centiles out of 100 repetitions of the classification experiment. (a) Performance of the 3-day heuristic. Each dot represents a different run out of 100 repetitions. Although straightforward to implement, the average performance of the heuristic does not provide the possibility of customizing the algorithm to specific needs, such as increasing the true-positive rate at the expense of a higher false-positive rate. AUC: area under the curve; ROC: receiver operating characteristic.





Feature	Times selected
Mean pre-breakfast readings	100
3-day heuristic pre-breakfast	100
Mean all post-prandial readings	93
Mean all post-brakfast readings	74
Mean all evening-meal readings	72
Count low pre-prandial readings	68
Mean all readings	61
Mean all post-breakfast readings	53
Age	49
Standard deviation of post-prandial readings	48
Count high pre-breakfast readings	46
Count high pre-lunch and pre-evening-meal readings	44
Risk factor: previous large-for-gestational-age baby	42
Minimum of pre-lunch readings	38

(c)



The ROC and AUC results of repeated experiments are shown in Figure 6. Figure 6 also shows the results of the lasso feature selection with the top 14 (top 10 in bold) and summarizes the number of times each was deemed relevant. Figure 6 summarizes the precision-recall curves. Results from the repetition of the validation step are described by the 5%, 50%, and 95% ROC values percentiles; these provide the median ROC curve and corresponding CIs, and the same is done for the precision-recall curves in Figure 6.

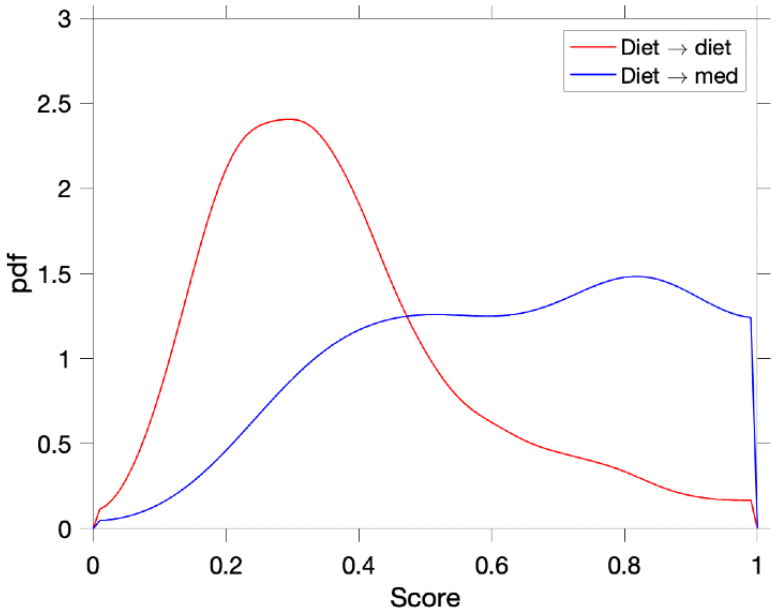
The histogram of AUC values drawn from the ROC curves is centered around a median value of 0.80, which supports the potential for clinical evaluation of the proposed algorithmic approach.

To evaluate the distribution of scores for a given model, we selected one close to the median performance shown in Figure 6, 0.80 AUC. The selected model operates with 4 features (mean of all BG values, mean of all prebreakfast BG values, max of all postbreakfast BG values, and raised 3-day prebreakfast). Figure 7 shows the distribution of scores for the 2 classes and demonstrates how the model can distribute scores for both classes, although an overlap is presented around the decision boundary. For completeness, additional figures showing the distribution of feature values are shown in the Multimedia

Appendix 1 for true positive, true negative, false positive, and false negative.

To compare performance with the standard of care, we evaluated the algorithm against the 3-day heuristic currently used in the trusts, considering that treatment should be considered when 3 or more consecutive BG readings of the same meal tag are over the designated threshold.

Figure 7. Distribution of scores from a model close to median performance (area under the curve 0.80). The model uses 4 features to perform its function. Although scores for the diet–diet group are clearly distributed on the left-hand side of the graph, the diet–med group presents higher variability.



As shown in Figure 6, the current heuristic performance is fixed to an area and is nonconfigurable. However, the proposed ML-based approach can make use of a different threshold (eg, to classify differently the score of 0 to 1 output by the logistic regression), thus allowing a change in the operating point of the system and allowing for a larger TPR at the expense of a slightly larger FPR. For example, from Figure 6, we could aim at 80% TPR by increasing our FPR by 20%, thus providing the possibility to the system to identify more women in need of medications, at the added cost of few women who will probably be screened by midwives and identified as false positives.

Discussion

The increased prevalence of GDM [6] and limited resources available to the NHS [7] pose a problem to the already burdened antenatal care services.

Predicting the need for pharmacological treatment could likely benefit women diagnosed with GDM by improving glycemic control, thereby leading to improved perinatal outcomes and avoiding complications such as large-for-gestational-age newborns or c-sections. Digital health technology such as GDm-Health can provide the real-time monitoring required to collect dense, longitudinal data sets and enable the delivery of clinical decisions quickly and efficiently to patients. Algorithms derived from real-world data obtained from GDm-Health could help midwives to optimize their clinical decision making and allow interventions, including medication, to be delivered earlier.

Conclusions

In this study, we have used ML on a large, anonymized data set from a population affected by GDM to design an algorithm capable of detecting the need for pharmacological treatment.

The strength of our study lies in the use of a large, multisite, real-world data set to validate our results. Predictors selected by our ML algorithm match most of the predictors included in the state of the art [11-16] and are enhanced by the use of risk factors and other demographic information available as part of routinely collected data by GDm-Health.

The logistic algorithm employed was experimentally tested against LightGBM and Random Forest algorithms. However, when applied to the same features and methodology, these comparator algorithms did not significantly improve AUC performance (both reporting a median of 0.81 AUC).

The aggregated results of the trained logistic regression models achieved an average AUC of 0.80, which is significant to justify future work to evaluate and validate this algorithm in real clinical settings.

Some of the limitations of this study are very common to other mHealth systems, including the challenge of user-reported data such as medication and BG data, which may be inaccurate or missing. However, in the case of GDm-Health data, user retention and user adherence have been very high, with only 4% of profiles being excluded because of complete disengagement with the system (117 women with no readings).

Given the longitudinal nature of this data set (ie, from 2014 to 2019) and the heterogeneous nature of each trust, women at each trust could have been subject to different clinical

management processes, adding to the complexity of the ML task.

Nevertheless, we demonstrate that our algorithm predicts the requirement for pharmacological therapy, and we show the superiority of our approach against a heuristic currently employed in clinical settings.

The very likely future introduction of ML algorithms to aid the work of health care professionals and to support patients coping with their conditions requires the validation of the technology using real-world data sets such as the one provided by GDm-Health. We intend to clinically validate the performance of the algorithm further by evaluating its real-time performance on a data set used for clinical operations. To that end, we will first pursue a posthoc analysis on a subset of data not used to design the algorithm and then deploy an implementation of the algorithm alongside GDm-Health to monitor its real-time performance (ie, predictions performed on a daily basis on updated BG daily readings) against decisions performed by health care professionals. Finally, repeated validation and postmarket evaluation strategies will be employed to continuously validate the algorithm against clinical decision making made by health care professionals.

Future work may include new analyses of the GDm-Health data set to include other variables that might identify a change in clinical patient management (eg, including the trust name as a predictor), considering variable lengths of predictive windows (eg, computing features at 2, 3, or 4 weeks before a medication event), or considering the problem as a time-to-event prediction (via Cox proportional hazards, etc).

Acknowledgments

The authors would like to thank all members of the TREAT-GDM research team, the patients who participated in the study, and those who had used the GDm-Health system. A special thanks to Ian Gallen for his invaluable contributions. The team acknowledges the support of Roberto Liddi, Rodrigo Jazinski, Michael Griffiths, Anna Muszkiewicz, James Farrant, Manoj Krishnamoorthy, and Darren Gibb.

Authors' Contributions

CV performed the analysis and wrote the first draft. LM, CV, LT, and DC contributed to the study methodology. All the authors contributed to revising and finalizing the paper.

Conflicts of Interest

Sensyne Health, plc, acquired the rights of the GDm-Health system from the University of Oxford, and it is the company that develops, maintains, and commercializes the GDm-Health system for the management of GDM. LM was supported by the National Institute for Health Research Oxford Biomedical Research Centre and works part time for Sensyne Health, plc. CV works for Sensyne Health, plc, and is a visiting lecturer in the Department of Engineering Science, University of Oxford. LT, DC, and PW work part time or consult for Sensyne Health, plc.

Multimedia Appendix 1
Error analysis of true positives, true negatives, false positives, and false negatives for a given selected model.
[DOCX File , 181 KB - jmir_v23i3e21435_app1.docx]

References

1. American Diabetes Association. Gestational diabetes mellitus. Diabetes Care 2003 Jan;26 Suppl 1:S103-S105. [doi: 10.2337/diacare.26.2007.s103] [Medline: 12502631]
2. Kim C. Maternal outcomes and follow-up after gestational diabetes mellitus. Diabet. Med 2014 Feb 16;31(3):292-301. [doi: 10.1111/dme.12382]
3. Fraser A, Lawlor DA. Long-term health outcomes in offspring born to women with diabetes in pregnancy. Curr Diab Rep 2014 Mar 25;14(5):489 [FREE Full text] [doi: 10.1007/s11892-014-0489-x] [Medline: 24664798]
4. Diabetes Association A. 13. Management of Diabetes in Pregnancy: Standards of Medical Care in Diabetes—2018. Dia Care 2017 Dec 08;41(Supplement 1):S137-S143. [doi: 10.2337/dc18-s013]
5. Walker JD. NICE guidance on diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period. NICE clinical guideline 63. London, March 2008. In: Diabet Med. Diabetes in pregnancy: management from preconception to the postnatal period Internet. National Institute for Health and Care Excellence (NICE); Sep 2008:1025-1027.
6. Farrar D, Simmonds M, Griffin S, Duarte A, Lawlor D, Sculpher M, et al. The identification and treatment of women with hyperglycaemia in pregnancy: an analysis of individual participant data, systematic reviews, meta-analyses and an economic evaluation. In: Health Technol Assess. Prevalence of gestational diabetes in the UK and Republic of Ireland: a systematic review Internet. NIHR Journals Library; Nov 2016:1-348.
7. Topol E. The Topol review: preparing the healthcare workforce to deliver the digital future. Internet. 2019. URL: https://topol.hee.nhs.uk/ [accessed 2021-02-20]

8. Agarwal S, Tamrat T, Glenton C, Lewin S, Henschke N, Maayan N, et al. Decision-support tools via mobile devices to improve quality of care in primary healthcare settings. Cochrane Effective Practice and Organisation of Care Group, editor. Cochrane Database Syst Rev Internet 2018 Feb 2018 Feb:19. [doi: 10.1002/14651858.cd012944]
9. Homko CJ, Deeb LC, Rohrbacher K, Mulla W, Mastrogiannis D, Gaughan J, et al. Impact of a telemedicine system with automated reminders on outcomes in women with gestational diabetes mellitus. Diabetes Technol Ther 2012 Jul;14(7):624-629 [FREE Full text] [doi: 10.1089/dia.2012.0010] [Medline: 22512287]
10. Mackillop L, Hirst JE, Bartlett KJ, Birks JS, Clifton L, Farmer AJ, et al. Comparing the Efficacy of a Mobile Phone-Based Blood Glucose Management System With Standard Clinic Care in Women With Gestational Diabetes: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Mar 20;6(3):e71. [doi: 10.2196/mhealth.9512]
11. Ali A, Shastry S, Nithiyananthan R, Ali A, Ganapathy R. Gestational diabetes–Predictors of response to treatment and obstetric outcome. European Journal of Obstetrics & Gynecology and Reproductive Biology 2018 Jan;220:57-60. [doi: 10.1016/j.ejogrb.2017.11.014]
12. Barnes RA, Wong T, Ross GP, Jalaludin BB, Wong VW, Smart CE, et al. A novel validated model for the prediction of insulin therapy initiation and adverse perinatal outcomes in women with gestational diabetes mellitus. Diabetologia 2016 Jul 8;59(11):2331-2338. [doi: 10.1007/s00125-016-4047-8]
13. Ducarme G, Desroys du Roure F, Grange J, Vital M, Le Thuaut A, Crespin‐Delcourt I. Predictive factors of subsequent insulin requirement for glycemic control during pregnancy at diagnosis of gestational diabetes mellitus. Int J Gynecol Obstet 2019 Jan 11;144(3):265-270. [doi: 10.1002/ijgo.12753]
14. Lobner K, Knopff A, Baumgarten A, Mollenhauer U, Marienfeld S, Garrido-Franco M, et al. Predictors of Postpartum Diabetes in Women With Gestational Diabetes Mellitus. Diabetes 2006 Feb 27;55(3):792-797. [doi: 10.2337/diabetes.55.03.06.db05-0746]
15. Meshel S, Schejter E, Harel T, Maslovitz S, Germez N, Elimelech B, et al. Can we predict the need for pharmacological treatment according to demographic and clinical characteristics in gestational diabetes? The Journal of Maternal-Fetal & Neonatal Medicine 2015 Aug 28;29(13):2062-2066. [doi: 10.3109/14767058.2015.1077225]
16. Watanabe M, Katayama A, Kagawa H, Ogawa D, Wada J. Risk Factors for the Requirement of Antenatal Insulin Treatment in Gestational Diabetes Mellitus. Journal of Diabetes Research 2016;2016:1-6. [doi: 10.1155/2016/9648798]
17. Information CO. Anonymisation: managing data protection risk code of practice. ICO 2012:2012. [doi: 10.5040/9781509922246.0021]
18. Duran A, Sáenz S, Torrejón MJ, Bordiú E, Del Valle L, Galindo M, et al. Introduction of IADPSG criteria for the screening and diagnosis of gestational diabetes mellitus results in improved pregnancy outcomes at a lower cost in a large cohort of pregnant women: the St. Carlos Gestational Diabetes Study. Diabetes Care 2014 Sep;37(9):2442-2450. [doi: 10.2337/dc14-0179] [Medline: 24947793]
19. Diabetes in pregnancy: Management of diabetes and its complications from pre-conception to the postnatal period. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/CG63 [accessed 2019-08-02]

Abbreviations

AUC: area under the receiver operating characteristic curve
BG: blood glucose
FPR: false-positive rate
GDM: gestational diabetes mellitus
mHealth: mobile health
ML: machine learning
NHS: National Health Service
OUH: Oxford University Hospitals
RBH: Royal Berkshire Hospitals
ROC: receiver operating characteristic curve
TPR: true-positive rate

Edited by G Eysenbach; submitted 15.06.20; peer-reviewed by Z Yin, C Lindvall; comments to author 19.09.20; revised version received 17.11.20; accepted 17.01.21; published 10.03.21.

Please cite as:
Velardo C, Clifton D, Hamblin S, Khan R, Tarassenko L, Mackillop L
Toward a Multivariate Prediction Model of Pharmacological Treatment for Women With Gestational Diabetes Mellitus: Algorithm Development and Validation
J Med Internet Res 2021;23(3):e21435
URL: https://www.jmir.org/2021/3/e21435
doi:10.2196/21435
PMID:33688832

Original Paper

The Perceived Impact and Usability of a Care Management and Coordination System in Delivering Services to Vulnerable Populations: Mixed Methods Study

Rubina Rizvi[1], PhD, MD; Courtney VanHouten[1], MA; Tiffani J Bright[1], PhD; Mollie M McKillop[1], PhD; Shira Alevy[1], MEd; David Brotman[1], MS; Megan Sands-Lincoln[1], MPH; Jane Snowdon[1], PhD; Barbie J Robinson[2], MPP, JD; Carolyn Staats[2], AB; Gretchen P Jackson[1], PhD, MD; William J Kassler[1], MPH, MD

[1]IBM Watson Health, Cambridge, MA, United States

[2]Department of Health Services, Sonoma County, Santa Rosa, CA, United States

Corresponding Author:
Rubina Rizvi, PhD, MD
IBM Watson Health
75 Binney St
Cambridge, MA, 02142
United States
Phone: 1 952 237 7660
Email: rubina.rizvi@ibm.com

Abstract

Background: People with complex needs, such as those experiencing homelessness, require concurrent, seamless support from multiple social service agencies. Sonoma County, California has one of the nation's largest homeless populations among largely suburban communities. To support client-centered care, the county deployed a Care Management and Coordination System (CMCS). This system comprised the Watson Care Manager (WCM), a front-end system, and Connect 360, which is an integrated data hub that aggregates information from various systems into a single client record.

Objective: The aim of this study is to evaluate the perceived impact and usability of WCM in delivering services to the homeless population in Sonoma County.

Methods: A mixed methods study was conducted to identify ways in which WCM helps to coordinate care. Interviews, observations, and surveys were conducted, and transcripts and field notes were thematically analyzed and directed by a grounded theory approach. Responses to the Technology Acceptance Model survey were analyzed.

Results: A total of 16 participants were interviewed, including WCM users (n=8) and department leadership members (n=8). In total, 3 interdisciplinary team meetings were observed, and 8 WCM users were surveyed. WCM provided a central shared platform where client-related, up-to-date, comprehensive, and reliable information from participating agencies was consolidated. Factors that facilitated WCM use were users' enthusiasm regarding the tool functionalities, scalability, and agency collaboration. Constraining factors included the suboptimal awareness of care delivery goals and functionality of the system among the community, sensitivities about data sharing and legal requirements, and constrained funding from government and nongovernment organizations. Overall, users found WCM to be a useful tool that was easy to use and helped to enhance performance.

Conclusions: WCM supports the delivery of care to individuals with complex needs. Integration of data and information in a CMCS can facilitate coordinated care. Future research should examine WCM and similar CMCSs in diverse populations and settings.

(J Med Internet Res 2021;23(3):e24122) doi:10.2196/24122

KEYWORDS

vulnerable population; managed care; data integration; advanced technologies; usability; mixed methods study

Introduction

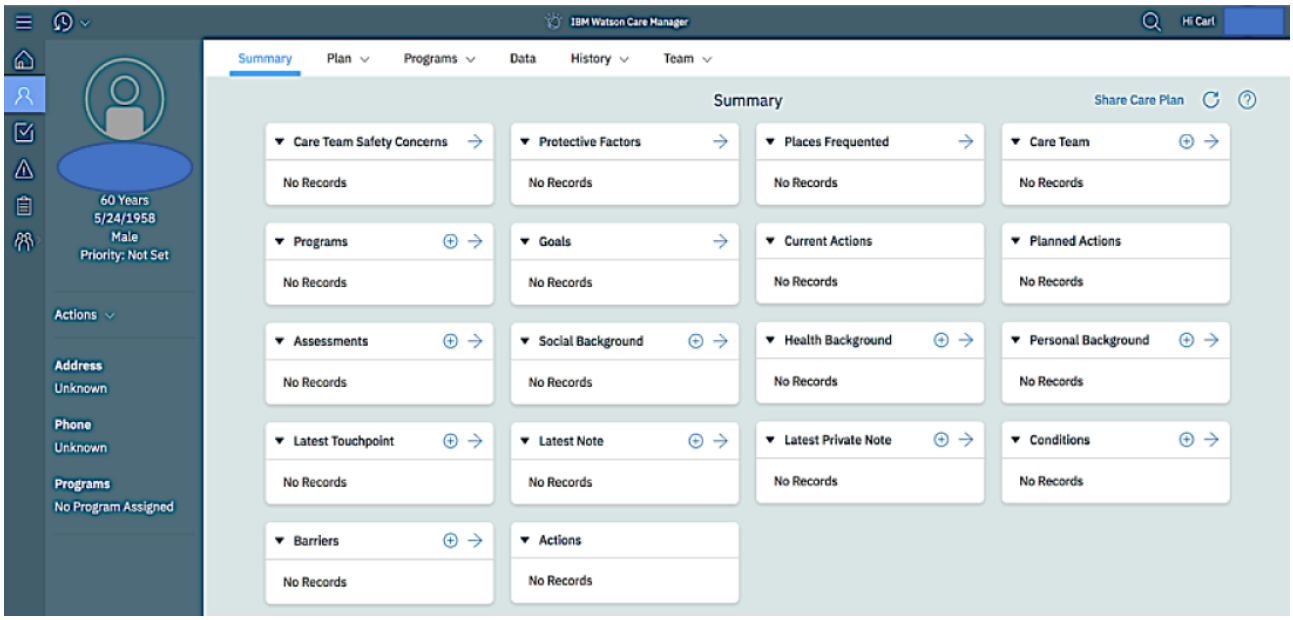
Background

Providing comprehensive services to vulnerable populations is a complex task requiring effective and efficient collaboration and resource alignment among various safety net agencies [1,2]. Care management processes are often complicated when such organizations operate in information silos. These processes often involve agencies such as health care, law, housing, and law enforcement, which tend to work independently and lack effective, seamless communication across agencies [3]. To mitigate this growing challenge, synchronization of agencies at different levels is required to successfully provide holistic, client-centered care in an effective and timely manner. These existing silos could be further addressed by using technology as a common platform, which could serve as a reliable information repository for data aggregation, reporting, and exchange. Currently, limited evidence exists about the use, benefits, and shortcomings of social management tools built on advanced technologies.

Sonoma County is a large county in California, with over 500,000 residents [4] and a disproportionately large homeless population compared with similarly sized communities [5]. The 2019 Sonoma County Homeless Census and Survey Comprehensive Report [6] suggested that approximately 3000 homeless people live primarily in vehicles (29%) and emergency shelters (25%); on streets (24%); and in various other shelters (22%), such as tents, transitional housing, or abandoned buildings [6]. This homeless crisis has been intensified by natural disasters, most recently the Kincaid Fire (October 23 to November 6, 2019), which threatened more than 90,000 structures and burned 77,758 acres, resulting in evacuations throughout Sonoma County [7]. This critical situation was recognized by the Sonoma County department leadership and increased the need for a solution that could help address the existing gaps by providing holistic, client-centered care for the county's residents with complex needs. In 2017, the board of supervisors established a 3-step approach for achieving their objectives by implementing the *Accessing Coordinated Care and Empowering Self-Sufficiency* (ACCESS) [8] initiative to identify the most vulnerable residents and provide them with coordinated services. In efforts to establish a rapid response, the interdepartmental multidisciplinary team (IMDT) was created to coordinate cross-departmental services. The IMDT mainly comprises personnel from each social service department and eligibility specialists, who were hands-on Watson Care Manager (WCM) users, and leaders and executives from programs and services participating in the ACCESS initiative, who engaged indirectly with the WCM system and its output but were not hands-on users. Sonoma County partnered with International Business Machines (IBM) to develop a Care Management and Coordination System (CMCS) to support the IMDT [9].

The primary objective of the CMCS is to successfully refer vulnerable residents to the services they need most and foster data sharing and collaboration among diverse care professionals to optimize service delivery. The CMCS tool consists of 2 components: Connect360 and WCM. Connect360 is an integrated data hub that receives data from participating agencies and generates a comprehensive, client-specific record as a single data source. These agencies include housing, human services, justice, child support services, substance use disorders, mental health, medical services, aging, and independence. WCM is the front-end interface (our study focus) that displays the consolidated client record to care providers. Using WCM, IMDT members actively enter, aggregate, and render up-to-date information, which enables them to develop integrated care plans for vulnerable clients (Figure 1) [3].

Figure 1. The front-end interface of Watson Care Manager.



Objectives

The purpose of this study is to evaluate the user-facing component of WCM to better understand its usability and perceived impact on service delivery processes at the Sonoma County Department of Health and Human Services (SC DHHS). We aimed to examine the role and perceived benefit of using WCM in case management, identify factors facilitating or impeding the use of WCM, and understand user acceptance of WCM with regard to perceived usability and ease of use of WCM.

Methods

Study Design

A mixed methods study, including observations, interviews, and surveys, was conducted at SC DHHS in December 2019. Purposive sampling was employed to recruit 2 sets of participants from SC DHHS: (1) WCM end users, which included eligibility specialists, and representatives from various government, law enforcement, and community agencies and (2) department leadership, which included individuals serving in an executive or management capacity.

Data Collection

The IMDT meetings were observed by four research team members (RR, CV, TB, and MS) with diverse backgrounds. The semistructured interview guides included questions focused on job role, initiation of client contact, client burden of care, client interaction and engagement, WCM use, and interview demographics. The interview guides used for WCM users and department leadership are included in Multimedia Appendices 1 and 2, respectively. Interviews were conducted on-site and over telephone; all were audio recorded, transcribed verbatim, and saved using a Health Insurance Portability and Accountability Act–compliant app Otter (version 2.1.17) [10]. Before each interview, the WCM users completed the Technology Acceptance Model (TAM) [11] survey (Multimedia Appendix 3) to assess the perceived usefulness (PU) and perceived ease of use (PEOU) of WCM.

Data Analysis

Data were analyzed using thematic analysis based on grounded theory [12]. Coding was performed sequentially on observations and interviews by 4 research team members (RR, CV, TB, and MM). One common transcript from each of the 3 data collection methods (ie, observations, interviews of WCM users, and department leadership interviews) was randomly selected and independently coded to generate an initial set of codes. Coders held 2 adjudication sessions to review codes and achieve consensus. Once a consensus was reached, a final code book for each transcript source, including interviews and observations, was created and applied to the remaining transcripts. Coding reliability was established through consensus among all coders. Dedoose (version 8.3.18) [13] was used to support qualitative data analysis.

Ethics Statement

All study participants gave written consent to participate in personal audio-recorded interviews. The study was deemed as exempt from human subjects research regulations by the Western Institutional Review Board.

Results

In total, 3 IMDT meetings were observed; each meeting lasted approximately 3 hours, and field notes were taken by each observer. The WCM tool was used during these meetings, and an example of the summary home page of the front-end interface is shown in Figure 1.

Interviews were conducted with 8 WCM users and 8 department leadership members. Of the 16 interviews, 13 were conducted in person and 3 were conducted over the telephone. Interviews with WCM users lasted for an average of 60 min; the longest interview lasted approximately 130 min. Department leadership interviews averaged approximately 40 min, whereas one leadership interview with 2 individuals lasted for 140 min. The study participant characteristics are aggregated and summarized in Table 1.

Table 1. Descriptive characteristics of Watson Care Manager users and department leadership (N=16).

Participants	Value, n (%)	Age (years), range	Age (years), mean (SD)	Gender (female), n (%)	Education (master's degree or higher), n (%)	Technology skills (intermediate skill level or higher), n (%)
WCM[a] users (n=8)	8 (50)	31-62	47 (13.2)	6 (75)	5 (63)	7 (88)
Department leadership (n=8)	8 (50)	42-58	51.6 (6.2)	5 (63)	6 (75)	N/A[b]

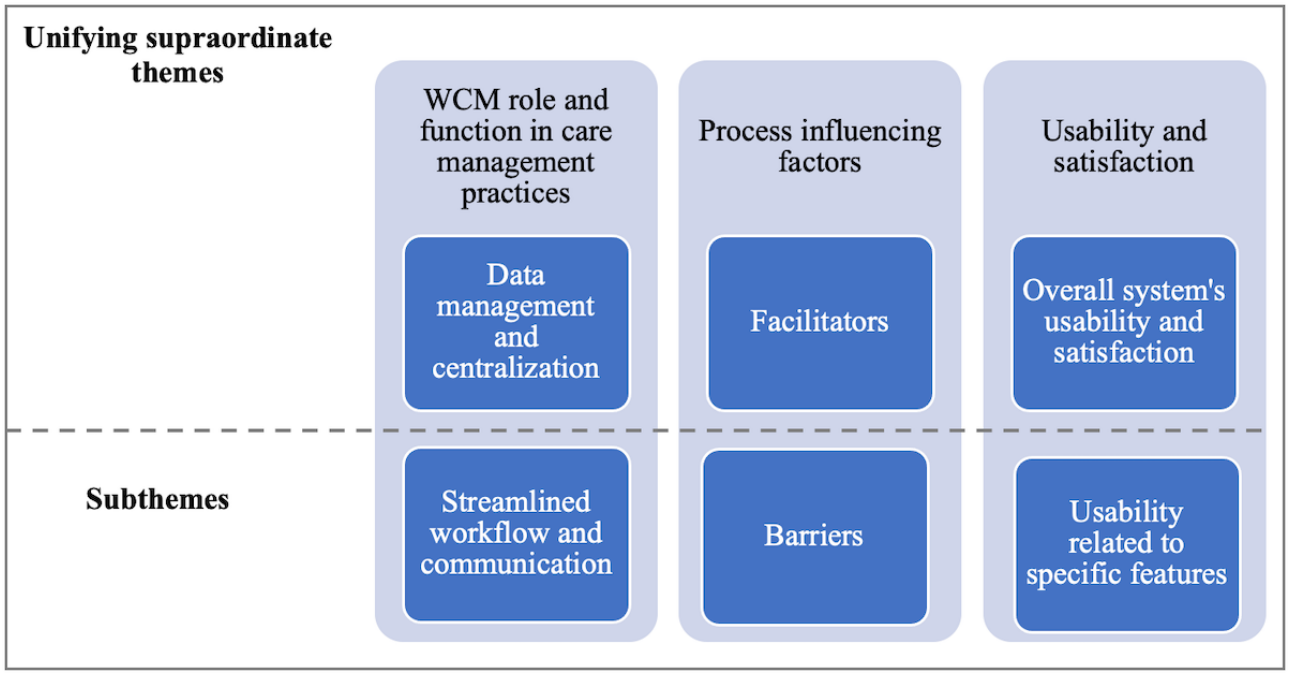
[a]WCM: Watson Care Manager.

[b]N/A: not applicable.

Analysis of the transcripts resulted in a list of codes. For reporting purposes, we included only the top 10 most frequently applied codes stemming from the analysis of 3 transcript sources (Multimedia Appendix 4).

The most granular-level codes were grouped as follows: (1) WCM role and function in care management practices with subthemes of data centralization and streamlined workflows (eg, shared decision making and information aggregation), (2) process influencing factors having subthemes of facilitators and barriers (eg, community demographics and data privacy), and (3) usability and satisfaction with subthemes overall usability and specific tool functionalities (eg, interaction with the interface and alerts; Figure 2).

Figure 2. The 3 unifying supraordinate themes and the corresponding subthemes. WCM: Watson Care Manager.



Unifying Theme: WCM Role and Function in Care Management Practices

Observations of IMDT meetings consistently suggested that WCM was used as the main tool for data review, both by leaders and users. WCM provided IMDT team members with an opportunity to have a collaborative team discussion about a client's status and review the data before the IMDT meeting. Despite its advantages, it was observed that 2 WCM users continued to reference paper copies of case information as a supplemental resource for facilitating updates and discussion.

Slight differences were observed between department leadership and WCM users. Leadership primarily discussed the positive impact WCM had on employees and department operations, whereas users tended to discuss the impact of WCM in the context of facilitating daily task accomplishments, establishing client rapport, and strengthening existing relationships. Despite these differences, these interviews revealed a praxis-based approach whereby descriptions of task achievement were used to illustrate how WCM supported larger goals. Analysis of all the observations and interviews revealed 2 principal themes of use: (1) data management and centralization and (2) streamlined agency communication.

Subtheme: Data Management and Centralization

Both program leaders and hands-on users repeatedly referenced the role of WCM as a data management tool to centralize client information and serve as a platform to reliably retrieve and store client information. In addition to progress-related updates, WCM helped users to stay up to date on short-term, practical tasks such as taking clients to probation or court hearings, delivering public transit vouchers, or keeping medical follow-up appointments. The following excerpt from the field notes validated how WCM facilitated scheduling a client's appointment, thereby enhancing task performance efficiency:

> *During a client presentation, the case manager requested a new medical appointment, and the nurse made the medical appointment instantly. The appointment was scheduled before the conversation was even over.* [Observation note]

WCM users routinely retrieved and documented case notes in up to 7 different systems simultaneously before working in WCM. Furthermore, documentation of client interactions in multiple systems was necessary to meet state and federal reporting requirements; however, they did not have a platform to store or share case-specific information such as care plans or client goals:

> *If Watson didn't provide that service [providing a platform across multiple systems], I'm doing what I've been doing that is, I have seven systems, and I work in those seven systems to gather information.* [Participant #6, user]

> *We also don't have a place where our care plan is currently living. And so Watson is the only place that the care plan actually lives. The living document. So, I think that is really important.* [Participant #6, user]

> *I use it to go in and see what one of the other team members have [already completed] when there is something that needs to be done. To make sure that anything that needs to be done for the benefit of the client is being done.* [Participant #4, user]

WCM not only helped in care delivery but also helped users maintain their client's enrollment in government support programs through an *alerts* feature. Department leadership noted that securing and maintaining such support is integral to helping a client achieve stability but that it can sometimes be a burden on client managers whose clients are often enrolled in several programs with different requirements and renewal terms:

> *Well, I kind of go back to what I was saying about the alerts. If I missed the alert, and I didn't reach out to the client and inform the client this is upcoming [and] this has to be done, or you will lose the benefit [and]to go through the process of trying to get that benefit reestablished is a lengthy process. So, it helps the client, and it helps me to know that I'm doing my job.* [Participant #4, user]

In addition, by aggregating client eligibility information, WCM coordinated multiple application processes, allowing the client to apply for multiple programs at one time. This coordination reduced the cognitive burden on employees and helped reduce unnecessary client effort by avoiding the need to constantly obtain required documentation or navigate multiple application procedures:

> *A lot of [applications and eligibility] interactions we use Watson Care Manager. We can like coordinate to get verification before my interview [with the client]. So, at the time of the interview begins or the time I meet the client like they can already have everything ready so it's like "here you go. Like I have it and it's more straightforward my end that I can approve them all one time and just collect the application.*" [Participant #2, user]

Department leadership interviews echoed the resulting benefits of WCM by rendering data that are streamlined and centralized, serving as a one-stop shop to learn about the client's current status and to plan out the next steps:

> *That's one of the things that everybody has a great help for with Watson, how we use that information, bringing our respective data together, and work in a more coordinated fashion.* [Participant #15, department leadership]

Subtheme: Streamlined Workflow and Communication

Streamlined workflow and communication emerged as common subthemes, from interviews of both leaders and hands-on users. WCM played an essential role in effective and efficient communication not only within the agency but also between other agencies. More specifically, agencies related to health care, housing, and justice were able to communicate across silos to address time-sensitive client needs. Examples included understanding a client's current health status and making timely health care appointments, ascertaining the availability of housing spaces based on individual needs, and verifying the parole status of the client. Observations noted up-to-date exchange of information between WCM users from different agencies, resulting in a comprehensive, timely care delivery plan that prioritizes clients' needs and avoids potential adverse repercussions:

> *Client released from jail to the crisis residential unit and needs mental health services but refuses. Group discussion around options, pathways – how to get client to seek mental health. Probation officer reminded the criticality of client's compliance with receiving mental health treatment since it was one of the conditions of his parole.* [Observation note]

During interviews, both WCM users and department leadership discussed how WCM use and IMDT meetings facilitated real-time discussions among the pertinent care providers from the representative agency, enabling more informed, shared, and efficient decision making. Both users and department leadership highlighted specific descriptions of how case managers and other client specialists were able to advocate for their clients. Users also shared their prior experiences before the introduction of WCM and described operating in silos and on different information technology systems. They described the benefit of provisioning a holistic, client-centered approach with a common shared platform with specific case examples:

> *We could all talk about how the client is doing within the shelter. And so the [parole officer and shelter manager] worked with her on a plan...and everything got better from there. We can have a shared understanding.* [Participant #6, user]

Effects on communication and coordination led to changes in workflow processes, and they were described by one user as follows:

> *A lot of [applications and eligibility] interactions we use Watson Care Manager. We can like coordinate to get verification before my interview [with the client]. So, at the time of the interview begins or the time I meet the client like they can already have everything ready so it's like, here you go. Like I have it and it's more straightforward my end that I can approve them all one time and just collect the application.* [Participant #2, user]

In addition to integrating data and alerts from multiple agencies, WCM facilitated cross-program collaboration by providing a shared vocabulary. This common language and shared context fostered mutual understanding among agencies:

> *I mean we've spent many hours talking about like really silly things but like things that are weirdly fundamental for, like, one government agency talking to another like, what does date of admission mean? What does it mean to the jail, what does it mean to behavioral health? Are we tracking the right date of admission?... oh my god what a headache to have a conversation but at the same time, like us, understanding that one person's definition of this thing is radically different from someone else's is important.* [Participant #6, user]

Unifying Theme: Process Influencing Factors

Facilitators and barriers influencing the implementation and use of WCM were primarily identified through participant interviews from both executives and the hands-on users of WCM. During interviews, factors that shaped engagement with WCM were described.

Subtheme: Facilitators Influencing WCM Implementation and Use

WCM users from the IMDT and department leadership discussed opportunities to expand WCM to additional groups of clients in other US counties, apart from Sonoma County. The

IMDT attendees acknowledged the benefits and value of WCM and provided examples of how performance improvement enabled client success.

> *...so, it helps the client, and it helps me to know that I'm doing my job.* [Participant #4, user]

> *It's just everything's there in your fingertips and you just might have to scroll down a little bit, but everything there...my success rate is so much more...ultimately that's what matters so like improving care, [I am] pretty satisfied with it.* [Participant #2, user]

Department leadership described their viewpoint on the role of WCM in facilitating a collaborative environment. It was valuable in helping their transformation toward a client-centered culture. They also discussed WCM's flexibility, which enabled diverse departmental participation and improved care coordination. Multiple agencies recognized the value of WCM's role in streamlining care delivery and facilitating efficiency. Finally, this environment indicated a strong and widespread commitment to collaboration and willingness to participate in WCM use:

> *Our ability to show that [WCM] is effective helps us advocate for and leverage your resources. If they see it as successful then they are willing to put more resources into it.* [Participant #15, department leadership]

> *This work that you're [WCM] doing, and what has been accomplished thus far with Watson - I think is groundbreaking, and really has great potential to go to scale.* [Participant #15, department leadership]

In addition, users and department leadership highlighted the unique aspects of Sonoma County that may have contributed to WCM's success in their client population. Notably, the manageable, medium-sized county with active community-based agency participation facilitated and maximized WCM's utility and resulting outcomes:

> *I think Sonoma is in a very good place in that we're medium sized county and we, we have a lot of community-based organizations that work with us are already under contract and so it's probably more possible in Sonoma County than anywhere to be successful.* [Participant #9, department leadership]

Subtheme: Barriers Influencing WCM Implementation and Use

Currently, there is a lack of awareness and knowledge about how the ACCESS initiative operates. Barriers exist and impede its widespread adoption across other counties within California:

> *...so I think we have work to do, and a lot of that will be about outreach!* [Participant #9, department leadership]

The concerns around data privacy and security and the complexity of legal requirements also present implementation challenges. Although leadership expressed the desire to participate in the ACCESS initiative, they expressed concerns about meeting strict federal and state reporting requirements on data sharing in addition to their struggle in identifying appropriate funding streams:

> *You know, we have these really, really, strict policies on our data, so we've been able to do some match where we've been able to get some cohort data and then match it against our system, but we haven't been able to into the universal system and you know that's been a little frustrating.* [Participant #9, department leadership]

In user interviews, participants discussed the possibilities of using WCM for simple data organization and commented that they had fewer data systems and necessary log-in requirements:

> *I have like nine systems...that I use on a frequent basis. And then I have to use the additional two or three, four programs...So, with that and then I have to use Watson Care Manager because there's not much integration with that so I have to do three different systems with my case notes, so you have three different IDs and passwords you have to log in and I have like 12.* [Participant #2, user]

One user, however, voiced concern about duplicate efforts and manual tasks required when WCM was not able to retrieve all client-related data from various resources. They suggested that improved integration was needed to make the processes more efficient:

> *So, while at the moment it isn't frustrating or feeling like an extra piece of work...it's just the system is not set up right now.* [Participant #3, user]

Unifying Theme: Usability and Satisfaction (Users' Perspective)

Users' perception of WCM was analyzed based on data from observations, user interviews, and surveys. Perspectives on usability, including both PU and ease of use, and satisfaction are presented in Table 2.

Table 2. Responses from the Technology Acceptance Model survey.

Variable and TAM[a] survey questions[b]	Reported TAM score, median (minimum-maximum)
PU[c]	
Enables me to accomplish tasks more quickly than other products	4 (3-7)
Improves my job performance	5 (3-7)
Increases my productivity	4.5 (2-7)
Enhances my effectiveness on the job	4.5 (3-7)
Makes it easier to do my job	4.5 (3-7)
I have found WCM[d] useful in my job	5 (3-7)
Median PU	5 (4-5)
PEOU[e]	
Learning to operate was easy	4.5 (2-7)
Easy to get to do what I want it to do	5 (3-6)
Interaction is clear and understandable	5 (3-7)
Flexible to interact with	5.5 (2-7)
Easy for me to become skillful	6 (2-7)
I found WCM easy to use	5 (2-7)
Median PEOU	5 (4.5-6)
Overall combined median	5 (4-6)

[a]TAM: Technology Acceptance Model.

[b]The Technology Acceptance Model version used in this study had 12 questions, 6 assessing perceived usefulness and 6 assessing perceived ease of use, and they were scored on a 7-point Likert scale where 1=extremely disagree and 7=extremely agree.

[c]PU: perceived usefulness.

[d]WCM: Watson Care Manager.

[e]PEOU: perceived ease of use.

Subtheme: Overall Perceived Usability and Satisfaction

Overall, WCM users found the system to be well suited for their needs, serving as a one-stop shop for users by providing them with client-related data under one coherent platform. Users described WCM as being able to reliably house client information and provide easy access to relevant, time-sensitive information. Most importantly, WCM did not simply replicate previously existing electronic or paper forms but dynamically supported client care delivery by providing access to the latest data on client status and progress-related updates. Interviews also indicated the agility of WCM in allowing for rapid customization and the creation of a tool personalized according to their preferences.

TAM surveys were completed by WCM users and, on average, took 10 to 15 min (Table 2).

User feedback also suggested that WCM was not being used to its full potential and that specific WCM features were underused. WCM users suggested that additional individualized training might allow for better optimization of WCM capabilities:

> *...it's almost like having a Ferrari, and only knowing how to drive a Ford.* [Participant #4, user]

Subtheme: Usability Related to Specific Features

The *alerts* feature was one characteristic of the tool that was valuable to WCM users. A benefits coordinator used the alerts feature to notify managers about programs for which their client was eligible, to set reminders for application deadlines, or to notify regarding benefits renewal dates. Users described the alerts as valuable reminders that helped them monitor and maintain their clients' enrollment and avoid coverage gaps. Department leadership interviews also expressed agreement on the benefits of the alert feature. Leadership noted that securing and maintaining such support is integral to helping a client achieve stability but is often a burden on client managers whose clients are often enrolled in several programs with different requirements and renewal terms:

> *Oh, yeah, I have one thing that I can put up as why Watson Care Managers are so needed is to share the information piece, and the alerts, put on that front page of the summary page.* [Participant #2, user]

Users also mentioned 2 other features that they thought were useful. Both the summary (with demographics) page and the availability of a clients' picture facilitated with their tasks by providing succinct, identifiable, and useful data:

> *The one thing that I think is the most useful is when you look at is the profile. The first thing is that demographics that picture. The name, the date of birth, the address and the phone number. That is my everything I minded the most useful.* [Participant #2, user]

Findings from observations, interviews, and the TAM survey indicated that users found WCM to be well suited for their needs. A high level of PU (PU=5) and PEOU (PEOU=5) was indicated as a result of the TAM survey, and WCM was considered a reliable and useful tool that was easy to use and helped improve participants' overall productivity.

It should be noted that some WCM users indicated concern about the lack of interface optimization, such as unnecessary clicking and scrolling. The few users that reported this were not familiar with the system, which influenced their ability to efficiently perform tasks. This feedback was further supported by the median TAM score of 4 (with a minimum-maximum value of 3-7), especially for the responses around *accomplishing tasks quicker*:

> *Part of it's about learning to use the system and part of it's about minimizing clicks...* [Participant #6, user]

Discussion

Principal Findings

Provision of effective, holistic, client-centered care for vulnerable populations presents a complex challenge requiring substantial collaboration and careful coordination. This study provides unique insights into how a technology solution serves as an essential anchoring tool in a case management model. It enables data consolidation across agencies under one platform and serves as a common communication channel. The tool cultivated an environment of shared decision making and strengthened relationships among agencies.

The CMCS data hub, comprising WCM and Connect360, helped to retrieve and consolidate data in one shared place. It provided case workers with most recent data, such as court dates, benefit eligibility due dates, and doctors' appointments for clients who often need timely actions. According to users, WCM played a valuable role in addressing the challenges unique to social work with homeless populations where information is often transient, including mobile phone numbers, addresses, and contact information. WCM users and department leadership also provided feedback on barriers, including lack of funding, complex reporting structures, and limited awareness about the ACCESS initiative. Despite these barriers, WCM was perceived as a useful and easy-to-use tool associated with high user satisfaction, as validated by observations, interviews, and TAM responses.

The WCM tool was recognized for its high level of usability and satisfaction. Key features included alerts, information summary page with demographics, the client picture, and the client list page, which served as an index to pull all the client records stratified by priority status. The overall TAM score results and the scores across the 2 dimensions of usefulness and ease of use reinforced findings from the qualitative analysis.

Furthermore, the TAM survey results may suggest a greater likelihood of future use of the WCM tool, consistent with previous literature [14]. No association was observed between the level of interaction, education, and self-perceived skills in the technology when comparing these variables with TAM results (Table 2), which is informative of its high usability as scored consistently across diverse sets of users.

Limitations

Although these findings illustrate the capabilities of an advanced technology-based tool (WCM) in enhancing the care delivery processes within a county's homeless population, the study is not without limitations. These findings focused only on one specific cohort of vulnerable, homeless individuals residing in a single county. We also did not objectively measure the impact of the tool on long-term client outcomes. Instead, the study examined the perceived and observed usability and effects on workflow from the perspectives of program leaders and hands-on users. We did not incorporate client perspectives in this study.

Conclusions

This mixed methods study provides a better understanding of the public health impact that social care management tools such as WCM may have on counties in need of care coordination. Future research investigating the impact of WCM on outcomes (eg, social, clinical, and economical) and the value of WCM for additional communities and diverse subpopulations is essential to build upon these existing findings.

Acknowledgments

The authors would like to acknowledge the work and dedication of Hannah Helmy, PhD; Linda Low; and Margaret Holly for their contribution to this research.

Conflicts of Interest

RR, CV, TB, MM, SA, DB, ML, JS, GJ, and WK are employees of IBM Watson Health.

Multimedia Appendix 1
Interview guide for Watson Care Manager users.
[PDF File (Adobe PDF File), 55 KB - jmir_v23i3e24122_app1.pdf]

Multimedia Appendix 2
Interview guide for department leadership.
[PDF File (Adobe PDF File), 39 KB - jmir_v23i3e24122_app2.pdf]

Multimedia Appendix 3
Technology Acceptance Model.
[PDF File (Adobe PDF File), 29 KB - jmir_v23i3e24122_app3.pdf]

Multimedia Appendix 4
Top 10 frequently applied codes stemmed from the analysis of 3 sources of transcripts. ACCESS: Accessing Coordinated Care and Empowering Self-Sufficiency WCM: Watson Care Manager WPC: Whole Person Care.
[PNG File , 700 KB - jmir_v23i3e24122_app4.png]

References

1. Committee on Quality of Health Care in America, Institute of Medicine. Crossing the Quality Chasm A New Health System for the 21st Century. Washington (DC): National academic press; 2001.
2. Bhatt J, Bathija P. Ensuring Access to Quality Health Care in Vulnerable Communities. Acad Med 2018 Sep;93(9):1271-1275 [FREE Full text] [doi: 10.1097/ACM.0000000000002254] [Medline: 29697433]
3. Snowdon JL, Robinson B, Staats C, Wolsey K, Strasheim T, Jackson GP, et al. Sonoma county empowering caseworkers to better serve the most vulnerable. Appl Clin Inform 2020 Aug;11(4):617-621.
4. United States Census Bureau: Sonoma County, California. United States Census Bureau. 2020. URL: https://data.census.gov/cedsci/profile?q=Sonoma%20County,%20California&g=0500000US06097&tid=ACSDP1Y2018.DP05 [accessed 2020-03-01]
5. Responding to Homelessness in Sonoma County. County of Sonoma, California. 2020. URL: https://sonomacounty.ca.gov/CAO/Homelessness/ [accessed 2020-03-01]
6. Sonoma County Homeless Survey and Consensus-A Comprehensive Report 2019. Applied Survey Research. 2019. URL: https://www.appliedsurveyresearch.org/homelessness-reports/2014/8/15/sonoma-county-homeless-census-and-survey [accessed 2020-03-01]
7. Kincade Fire. Wikipedia. 2019. URL: https://en.wikipedia.org/wiki/Kincade_Fire [accessed 2020-03-01]
8. Robinson B, Staats C. Access Sonoma: enabling multi-discipline teams to deliver safety net services to county residents with complex needs. 2019 Feb 15 Presented at: Think 2019; February 12-15, 2019; San Francisco, CA.
9. Snowdon JL, Robinson B, Staats C, Wolsey K, Sands-Lincoln M, Strasheim T, et al. Empowering caseworkers to better serve the most vulnerable with a cloud-based care management solution. Appl Clin Inform 2020 Aug;11(4):617-621. [doi: 10.1055/s-0040-1715894] [Medline: 32969000]
10. Otter Voice Meeting Notes. Apple App Store. URL: https://apps.apple.com/us/app/otter-voice-meeting-notes/id1276437113 [accessed 2019-12-01]
11. Lewis J. Comparison of Four TAM Item Formatsffect of Response Option Labels and Order. J Usability Stud 2019;14(4):E.
12. Corbin J, Strauss A. Basics of Qualitative Research Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: SAGE Publications; 2008.
13. Dedoose. Great Research. Made Easy!. Dedoose. URL: https://www.dedoose.com [accessed 2019-12-01]
14. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]

Abbreviations

ACCESS: Accessing Coordinated Care and Empowering Self-Sufficiency
CMCS: Care Management and Coordination System
IBM: International Business Machines
IMDT: interdepartmental multidisciplinary team
PEOU: perceived ease of use
PU: perceived usefulness
SC DHHS: Sonoma County Department of Health and Human Services
TAM: Technology Acceptance Model
WCM: Watson Care Manager

Edited by G Eysenbach; submitted 04.09.20; peer-reviewed by M von Wagner, A Angerer; comments to author 10.12.20; revised version received 28.12.20; accepted 13.01.21; published 12.03.21.

Please cite as:
Rizvi R, VanHouten C, Bright TJ, McKillop MM, Alevy S, Brotman D, Sands-Lincoln M, Snowdon J, Robinson BJ, Staats C, Jackson GP, Kassler WJ
The Perceived Impact and Usability of a Care Management and Coordination System in Delivering Services to Vulnerable Populations: Mixed Methods Study
J Med Internet Res 2021;23(3):e24122
URL: https://www.jmir.org/2021/3/e24122
doi:10.2196/24122
PMID:33709928

Original Paper

Key Variables for Effective eHealth Designs for Individuals With and Without Mental Health Disorders: 2^12-4 Fractional Factorial Experiment

Armando J Rotondi[1,2,3], DPhil[‡]; Jonathan Grady[4], DPhil; Barbara H Hanusa[1], DPhil; Gretchen L Haas[1,5], DPhil; Michael R Spring[6], DPhil; Kaleab Z Abebe[7], DPhil; James Luther[1], MA; John Gurklis[8], MD

[1]Mental Illness Research, Education and Clinical Center (MIRECC), VA Pittsburgh Health Care System, Department of Veterans Affairs, Pittsburgh, PA, United States

[2]Center for Health Equity Research and Promotion (CHERP), VA Pittsburgh Health Care System, Department of Veterans Affairs, Pittsburgh, PA, United States

[3]University of Pittsburgh Center for Behavioral Health, Media, and Technology, University of Pittsburgh, Pittsburgh, PA, United States

[4]Computer Science, Thomas College, Waterville, ME, United States

[5]Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[6]Department of Information Sciences and Technology, School of Information Sciences, University of Pittsburgh, Pittsburgh, PA, United States

[7]Center for Research on Health Care Data Center, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[8]Behavioral Health, VA Pittsburgh Health Care System, Department of Veterans Affairs, Pittsburgh, PA, United States

[‡]Armando Rotondi

Corresponding Author:
Armando J Rotondi, DPhil
Mental Illness Research, Education and Clinical Center (MIRECC)
VA Pittsburgh Health Care System
Department of Veterans Affairs
Research Office Bld, (151R-U)
University Drive C
Pittsburgh, PA, 15240
United States
Phone: 1 412 360 2369
Email: armandorotondi1@gmail.com

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/7/e31553

Abstract

Background: eHealth applications not only offer the potential to increase service convenience and responsiveness but also expand the ability to tailor services to improve relevance, engagement, and use. To achieve these goals, it is critical that the designs are intuitive. Limited research exists on designs that work for those with a severe mental illness (SMI), many of whom have difficulty traveling for treatments, reject or infrequently seek treatment, and tend to discontinue treatments for significant periods.

Objective: This study aims to evaluate the influence of 12 design variables (eg, navigational depth, reading level, and use of navigational lists) on the usability of eHealth application websites for those with and without SMI.

Methods: A 2^{12-4} fractional factorial experiment was used to specify the designs of 256 eHealth websites. This approach systematically varied the 12 design variables. The final destination contents of all websites were identical, and only the designs of the navigational pages varied. The 12 design elements were manipulated systematically to allow the assessment of combinations of design elements rather than only one element at a time. Of the 256 websites, participants (n=222) sought the same information on 8 randomly selected websites. Mixed effect regressions, which accounted for the dependency of the 8 observations within participants, were used to test for main effects and interactions on the ability and time to find information. Classification and

regression tree analyses were used to identify effects among the 12 variables on participants' abilities to locate information, for the sample overall and each of the 3 diagnostic groups of participants (schizophrenia spectrum disorder [SSD], other mental illnesses, and no mental illness).

Results: The best and worst designs were identified for each of these 4 groups. The depth of a website's navigation, that is, the number of screens users needed to navigate to find the desired content, had the greatest influence on usability (ability to find information) and efficiency (time to find information). The worst performing designs for those with SSD had a 9% success rate, and the best had a 51% success rate: the navigational designs made a 42% difference in usability. For the group with other mental illnesses, the design made a 50% difference, and for those with no mental illness, a 55% difference was observed. The designs with the highest usability had several key design similarities, as did those with the poorest usability.

Conclusions: It is possible to identify evidence-based strategies for designing eHealth applications that result in significantly better performance. These improvements in design benefit all users. For those with SSD or other SMIs, there are designs that are highly effective. Both the best and worst designs have key similarities but vary in some characteristics.

(J Med Internet Res 2021;23(3):e23137) doi:10.2196/23137

KEYWORDS

schizophrenia; severe mental illness; eHealth; eHealth design; website; usability; website design; website usability; fractional factorial design

Introduction

Background

Schizophrenia spectrum disorder (SSD) and other severe mental illnesses (SMIs) are commonly chronic, require ongoing treatment and support to obtain optimal health, and reduce the occurrence of symptom relapses and hospitalizations [1]. Unfortunately, treatments may not be sought, disengagement from mental health services is common, and individuals' impressions of these services are not always favorable. Approximately 50% of those with a SMI do not seek mental health treatment in any given year [2], and 75%-85% discontinue their antipsychotic medications for a significant period during any 2-year period [3]. In addition to medications, psychosocial services (eg, case management, family psychoeducation, and cognitive behavioral therapy) are key components of evidence-based treatment for those with an SMI because they improve well-being over and above medications alone [1,4]. However, for many, there is a lack of availability of evidence-based and effective psychosocial services [5]. In addition, and similar to medication use, noninitiation, poor adherence, and discontinuation are common [5]. Unfortunately, not receiving psychosocial services and/or medications is associated with poorer outcomes, including increased risk of relapse, hospitalization, suicide, and homelessness [1,6]. In one study, 49% of those with an SMI did not seek treatment in the past year, yet said they had a problem needing professional help [2]. When individuals with a mental illness were asked why they disengaged from treatment, they cited not being listened to, unsympathetic providers, lack of participation in decisions, and being dissatisfied with services [7]. Services that can improve availability, can meet service users where they are in terms of their needs and priorities, can be tailored to their individual preferences, can be integrated with their lifestyles, and can lead to improved patient-focused outcomes are needed.

The use of eHealth technologies is an innovative approach for creating delivery models that can achieve these goals. They have the potential to support services that are more widely accessible, convenient, personalized, and able to adapt to the needs of individual users. Unfortunately, although there are many examples of success [8,9], the use of eHealth technologies has not always been successful. Dr Eysenbach advanced the *Law of Attrition (Or: Why Do eHealth Users Discontinue Usage?)* [10]. In one of his cited examples, only 1% (12/1161) of enrollees completed a 12-week panic disorder web program. In another example, an *open* evaluation of MoodGym, a proven-to-be-effective web-based intervention for depression, 99.5% of participants discontinued before completing the 5 modules. Although initial interest on the part of participants for the interventions was high, enthusiasm was lost for several reasons, including that the websites were too complex, not intuitive, and had poor usability. The Law of Attrition by Dr Eysenbach highlights the critical need to design highly usable, intuitive eHealth interventions that are accessible even to those with low technical and reading skills. Unfortunately, for those with SMIs, who may have special cognitive needs, when eHealth applications are created using design models for the general public, the designs are ineffective, often to the point of being unusable [11,12].

Dearth of Empirical Basis for Designing eHealth Services

Due to insufficient research, there is a critical gap in the knowledge base needed to design eHealth technologies for individuals with SMI and those with cognitive impairments, special cognitive needs, and/or low technology skills [13]. Some sources for guidance do exist. A number of design guidelines are inclusive of conditions relevant to SMI [13,14]. Commonly, guideline recommendations are meant to be general enough to apply across several health conditions and illnesses. As a consequence, the recommendations can vary among guidelines, even for a given population or group. For example, a synthesis of recommendations from 20 existing guidelines [14], all of which addressed individuals with conditions that included cognitive impairments, identified only 3 recommendations that were endorsed by more than 50% of the guidelines. These recommendations, starting with the most frequently endorsed, were to use pictures, icons, and symbols with the text; use clear and simple text; and use consistent navigation and design on

every page. Only the latter 2 recommendations are consistent with the empirical base with SMI [11,12,15-18], although the specifics of how to achieve these were not described. Another common limitation is that because a guideline may be created to cover a broad range of conditions, the recommendations may be more appropriate for some than others. In the previous example, the first and most commonly endorsed recommendation has been found to be a poor design feature for those with SSD and others with SMI [19,20].

There have also been recommendations based on empirical usability studies. There is consistency in the findings for some of these recommendations and inconsistency with others. Currently, consistent recommendations include the use of *large navigation buttons* [11,15,17,18]; *text at a low reading level*, preferably fourth to fifth grade [11,12,15,18,20]; *a shallow navigational hierarchy* [11,12,16-18,20]; *explicit or concrete wording* of hyperlinks, labels, and headings [11,12,17]; *pop-up menus* that appear via cursor hovering, to aid navigation [11,20]; and the *language of intended users* [12,16].

Recommendations that have been inconsistent include the font size, of which some have found it is better to increase the size over standard and use larger font [15,18,21], whereas others have found using smaller font works best [20]. In this study, we report that the effect of font size can be dependent on the design environment and on whether the desired goal is to improve the ability to find information or to reduce the time it takes to find information, to use designs that are more exciting, and to be able to grab users' attention [22], whereas others have found the opposite, that plain designs with little distracting and superfluous content, images, or displays are best [19,20]. What is most needed to improve eHealth designs for those with SMI, as well as others, is evidence-based recommendations derived from empirical usability investigations that collect valid quantitative performance data. As the empirical foundations expand, additional recommendations will emerge, and current inconsistencies will be resolved.

Objective

We have conducted a research program focused on identifying the design needs of individuals with SMI [12,17,20,23], as have others [11,15,18,24]. On the basis of this evidence-base, we selected a set of 12 eHealth screen design variables with the potential to have the largest effects on usability for individuals with SMI. The aim of this study is to examine the relative influence of these 12 variables on the ability of users to navigate websites. To accomplish this, we algorithmically specified the designs and then created 256 websites for testing. These websites systematically varied the 12 design variables via a 2^{12-4} fractional factorial design [25], enabling an equivalent representation of the 12 design variables in testing combinations of the design. The final destination pages of all websites were identical; only the designs of the navigational screens and pathways were different. For analyses, participants were divided into 3 diagnostic groups: those with a SSD diagnosis, those with any other mental health diagnosis, and those with no mental health diagnosis.

Methods

Participant Recruitment and Training

Participants were recruited via convenience sampling from the Department of Veterans Affairs (VA) Pittsburgh Health Care System and non-VA community mental health outpatient treatment centers. The enrollment criteria were as follows: aged at least 18 years, physical ability to read the screen of a computer and use a mouse, and ability to read at a fifth-grade level. There were no requirements for prior computer, mouse, internet, or website use. This study was approved by the VA Pittsburgh Healthcare System Institutional Review Board.

To screen for reading ability, participants completed the reading subtest of the Wide Range Achievement Test. Data were collected over 3 separate sessions. In the first session, demographic data, questions about past computer use, and the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) were administered to all participants to assist in determining the presence of DSM-IV Axis I mental disorders [26]. One of the questions was to self-rate one's level of computer understanding using a 5-point scale. In the analyses, this was used as a proxy for how savvy participants were with technology. To simplify the analyses, this was collapsed into 3 levels of self-rated expertise: responses 1 and 2 represented *low*, 3 was *moderate*, and 4 and 5 were *high*. In the second session, a neurocognitive functioning measurement battery was administered to evaluate basic cognitive abilities. In the third session, participants were tasked to find information on 8 different websites. To ensure that all participants had the basic skills needed to navigate the websites, each was taken through a brief tutorial (covering topics such as mouse use, hyperlink text, pop-up menus, and page scrolling) [12]. All individuals who met the eligibility criteria were able to master these basic skills. Following this, the performance of the tested websites was evaluated.

Design of the Websites and Selection of the 12 Variables

In total, 12 interface design variables (Table 1), each at 2 levels, were used to algorithmically generate a full factorial design of 2^{12} (ie, 4096) website designs. From these 4096 website designs, we identified an algorithmically derived *fractional* set of 256 websites to create. That is, the design of each of the 256 websites was specified via a sequential 2^{12-4} fractional factorial [25], which is a common industrial engineering experimental design. We have been conducting a research program focused on identifying the eHealth design needs of individuals with SMI and have included in this program a focus on SSD. This research program created an eHealth design model for those with SMI, termed the flat explicit design model (FEDM) [12], which was initially composed of 6 design variables. As the program progressed, the FEDM grew to 18 design variables [17]. The 12 variables for this study were chosen to be the most influential design variables, based on a literature and internet review of design recommendations and our experiences creating and evaluating eHealth applications for those with SMI [27], and to be the most important from the 18 variables of the FEDM. The reading level of website content was assessed using the

Flesch-Kinkaid reading scale (as assessed by Microsoft Word) [28]. The contents (ie, topics and final destination articles) were taken from a previously created web-based intervention [27] and were identical in all websites; only the navigational designs of websites differed.

Table 1. The 12 website design variables.

Wording used for design feature	Design feature	Levels of dimensions in the factorial design	
		Low	High
Navigational depth	Number of pages one needs to navigate to get from the home page to a desired piece of information.	≤3	≥5
Number of hyperlinks on a page	Each hyperlink is counted as one link. They may be embedded in standard text (ie, nonhyperlinked text) or stand alone, buttons (images, icons, and logos) or tool bars, and pull-down/pop-up menus. This variable is the total number of hyperlinks on a page.	≤7	≥14
Pop-up menu	Use of a pop-up menu to display the hyperlinks on the page.	0	1
Reading level	The grade level of links, labels, and text.	≤7	≥9
Words per page	Number of words on a page.	<100	≥200
Screen length	Screen length of navigation pages (paging vs scrolling)	≤1	>1
Number of distinct navigation areas	The number of separate areas on a page where users will find hyperlinks that can navigate the site.	≤2	≥4
Font size of body text	The size of the text used in the website	10 point	13 point
Number of words per hyperlink	The number of words used in the hyperlinks to the application's contents.	≤3	≥6
Number of nonhyperlink graphic elements on a page	This includes images, pictures, graphics (eg, color bars), and figures on a page that are not hyperlinks. They are for *decoration* or illustration only.	0	≥3
Constant navigational toolbar	Navigational tool bars that are in the same area across all pages on a website.	0	1
Number of topic areas on a page	A topic is an area devoted to one subject, purpose, or theme. For example, it could be a welcome paragraph; a list of links to main topics in the website (with or without introductory text); a list of links to news about the latest research on a topic; or a place to enter data, such as a search engine box.	≤4	≥6

Website Design and Evaluation Procedures

Each subject was asked to find information on 6 specified topics, on each of the 8 websites. Our previous experience indicated that this level of effort would not overly tax anyone who met our selection criteria. Given that we did not want fatigue to influence the results during testing, we chose a relatively conservative number. Example tasks were to find information on *what causes schizophrenia*, *how schizophrenia is treated*, and *the side effects of medications used to treat schizophrenia*. Two steps were taken to eliminate a *learning* effect influencing the results across the sample of participants. The order of the 6 tasks and the order of the 8 websites varied from individual to individual. The task order for each website was assigned using randomly permuted blocks of size 6. To vary the order of testing of the websites across participants, the order was assigned using randomly permuted blocks of 8. Testing occurred in a research office. If a subject selected an item as his or her choice but the item was not the target, he or she was informed and instructed to continue searching for the item until the allotted task time had expired (3 minutes). Tasks were timed unobtrusively using a browser plugin that we had developed. When the time allotted to complete a task expired, the computer screen automatically went blank, and the next task was initiated. Recording of the website usability testing was accomplished using applications that we developed to operate with the Firefox internet browser. To minimize testing anxiety, each participant was read a script explaining that the procedures were to evaluate the websites and not the participant and that there were no right or wrong answers.

Participants' abilities to locate the requested information (ie, design effectiveness) and time to accomplish tasks (ie, design efficiency) were analyzed to identify the design variables and combinations of variables that created more or less usable websites (ie, facilitate or inhibit use). The maximum time allowed to look for information was 180 seconds per task. To eliminate the over dispersion of times to success, the time data were analyzed using the natural logarithms of seconds to find.

Statistical Analyses

Comparisons of baseline demographics across the 3 mental health groups were completed with analyses of variance for continuous measures and chi-square for categorical measures, with exact tests as needed for small samples. Percentages of success and times to success were averaged over each website. Regression tree analyses were used to identify patterns of the 12 dimensions that led to more or fewer successes at finding

the required information or more or less time to successfully find the information.

Comparisons of performance across each of the 12 website dimensions were performed first on the entire sample (n=222 participants for ability to find information and n=218 for time to find information because 4 participants did not find any of the targeted information) and then for each of the 3 diagnostic groups separately, using classification and regression trees (CART) for success (CART Salford Predictive Modeler v8.2) [29]. The main effects were tested for all dimensions, and the interactions indicated as potentially significant by CART were tested with mixed effect regression models. Nesting of website performance within an individual was included in the regressions.

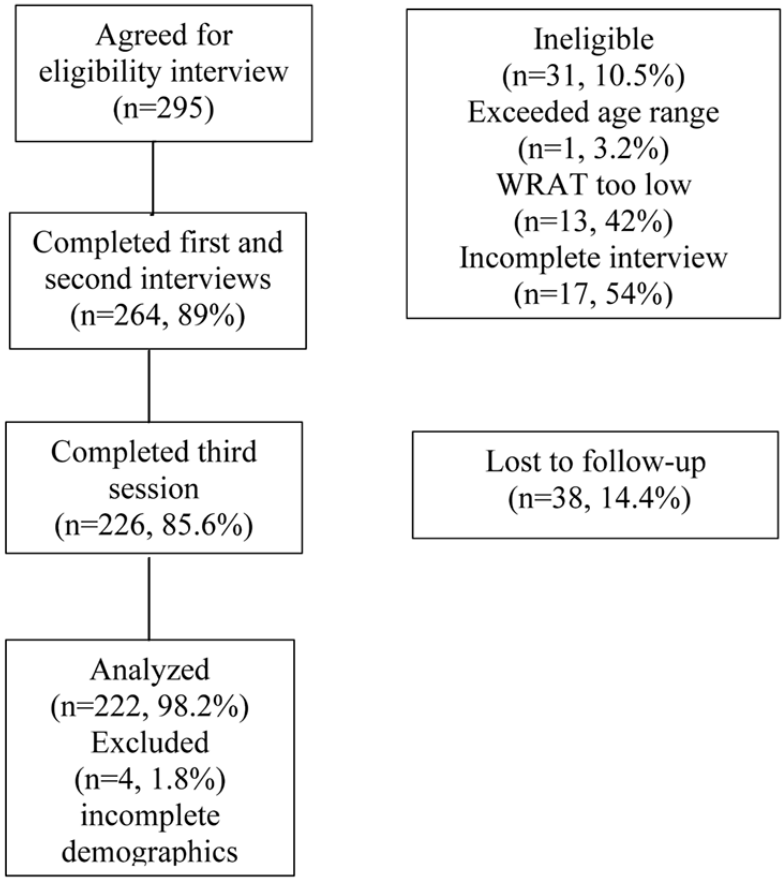
Regression tree procedures identify, among the predictor variables, the most efficient variable for splitting the observations into 2 groups. For example, it selected the 1 website dimension that best split higher performing websites (in terms of number of tasks completed successfully) versus lower performing websites. After the first split, each branch (descendent group) was split by the most efficient predictor variable within that diagnostic group. This created another level of branching. The amount of branching that occurs can be determined by specifying the number of observations in a descendent group that stops the branching or the number of branches desired. In this study, the tree was based on the percentage of tasks within a website that were completed correctly, and the branching was stopped with <12 person observations per website. As regression tree analyses do not account for multiple observations per individual subject, mixed effect regressions that can account for multiple assessments evaluated by the same individual were performed. The results of the regression tree analyses were used to identify possible interactions that would be tested in a multilevel model. Initial regression analyses included the 12 dimensions and interactions identified in the regression trees as well as the diagnostic variables of the 3 groups. Backward stepped regressions were used to remove the nonsignificant main effects and interactions. These models also indicated that the differences between the 2 non-SSD groups were not significant at the main effect or interaction effect level, and these 2 groups were combined and served as the baseline diagnostic group.

Results

Sample Description

A total of 295 participants met the eligibility criteria (Figure 1). Of these 295 participants, 264 (89.5%) completed the first and second data collection sessions for the study, and 226 (85.6%) completed the third session where the performance of the websites was assessed. Of the 226 participants, 4 were missing key demographic and/or computer experience information, resulting in 222 (75.2%) of the original 295 participants for this report. On the basis of mental health diagnoses, the 222 participants were placed in 1 of 3 diagnostic groups: no mental illness (83/222, 37.4%), mental illness without evidence of psychotic features (eg, depression and anxiety; 60/222, 27.0%), and SSD (79/222, 35.6%).

Figure 1. Consolidated Standards of Reporting Trials chart. WRAT: Wide-Range Achievement Test.



The 222 participants ranged in age from 23 to 75 years (mean 52.2, SD 9.2), and African Americans comprised 49.5% (110/222) of the sample (Table 2). Those with SSD, when compared with the other 2 groups, were more likely to be male, less likely to be employed, and more likely to be retired or disabled. In terms of technology, individuals with SSD were less likely to have a computer available at home, more likely to self-assess as having little or no computer understanding, and less likely to self-assess as having high computer understanding.

Table 2. Demographic and clinical characteristics split by diagnostic group (N=222).

Demographic variable	SSD[a] (n=83)	Mental illness other than SSD (n=60)	No current mental health diagnoses (n=79)	Comparison of the 3 diagnostic groups (*P* value)
Age (years)[b], mean (SD); range	53.1 (8.7); 29-71	52.1 (8.6); 23-72	51.6 (10.2); 23-75	.58
Gender, n (%)[c]				.02
Male	69 (84)	45 (75)	50 (63)	
Female	14 (16)	15 (25)	29 (37)	
Race, n (%)[d]				.11
White	46 (58)	24 (40)	49 (51)	
African American	35 (42)	36 (60)	39 (39)	
Education level, n (%)[e]				.14
Less than or equal to high school graduate	39 (47)	16 (27)	31 (39)	
Post high school training	35 (42)	33 (55)	34 (43)	
College degree or more	19 (9)	11 (18)	14 (18)	
Employment status, n (%)[f]				<.001
Full or part time	15 (18)	23 (38)	39 (49)	
Retired/disability only	68 (82)	37 (62)	40 (51)	
Self-reported computer understanding, n (%)[g]				.06
None/little	31 (38)	11 (18)	19 (24)	
Some	28 (34)	22 (37)	25 (32)	
Good/complete	23 (28)	27 (44)	34 (44)	
Computers available at home, n (%)[h]	28 (34)	34 (57)	44 (56)	.006

[a]SSD: schizophrenia spectrum disorder.

[b]F_{2221}=0.55.

[c]χ^2_2=8.3.

[d]χ^2_4=4.4.

[e]χ^2_4=6.9.

[f]χ^2_2=18.0.

[g]χ^2_4=9.0.

[h]χ^2_4=10.3.

Website Usability

All Participants Pooled

When all participants for the 256 websites were pooled, the average rate of success at finding information was 33.2%. The mean time to correctly find information was 49.0 seconds (median 40, range 7-180). The SSD group, when separately compared with the other 2 groups (ie, mental illness other than SSD and no mental illness), had a lower success rate (*P*=.007 for each) and took longer to find information (*P*=.001 for each; Table 3). All the 256 websites that were used in this study were created specifically for the study. Consequently, no participant had any experience with the websites they were tested on and were seeing and using each website for the first time.

Table 3. Overall performance (N=222).

Ability to find information	SSD[a], n=83		Mental illness other than SSD, n=60		No current mental health diagnoses, n=79		Significance
	Mean (SD)	Median	Mean (SD)	Median	Mean (SD)	Median	
Success rate at finding information per website (%)	28.6 (29.4)	16.7	37.8 (30.9)	33.3	34.5 (30.3)	33.3	Median, P=.007[b]
Time to find information per website (s)	59.8 (38.0)	51.1 (8-180)	44.4 (31.1)	35.3 (7.3-180)	42.8 (30.9)	35.3 (9.5-180)	Mean, P<.001[b]

[a]SSD: schizophrenia spectrum disorder.

[b]Significance levels based on mixed models that accounted for nesting of websites within participants.

Which Design Variables Influence Ability to Find Information: Mixed Regression Models, All Participants

Mixed regression models were created to identify the variables with main effects and interaction effects on the ability of participants to successfully find information. One of the models (Table 4, model 1) included the 12 design dimensions (Table 2) and the 3 diagnostic groups. In this model, the higher level of each of the 6 design variables leads to worse performance: higher navigational depth (≥5), higher number of hyperlinks on a page (≥14), higher reading level (≥ninth grade), larger font size (13 points), longer than 1 screen length of contents (ie, contents that require scrolling to see all of it), and higher number of navigation areas on a page (≥4); however, the higher level of 1 design variable use of a pop-up menu to display hyperlinks on a page improved performance. Having an SSD had a negative influence on the ability to find information (although it must be noted that this occurred within the 3-minute time limit imposed on finding a given piece of information). There were no significant differences between the other 2 diagnostic groups, that is, the difference was between SSD and the other 2 groups combined. A second mixed regression model (Table 4, model 2) was developed that included the self-rated measure of participant computer understanding. With this variable in the model, SSD (vs others) was no longer a significant variable.

Table 4. Multivariable models of percentage of tasks completed successfully.

Variables	Coefficient	95% CI	Significance (*P* value)
Model 1: Design dimensions and diagnostic groups			
Dimensions			
Navigational depth	−0.318	−0.348 to −0.289	<.001
Hyperlinks per page	−0.083	−0.108 to −0.060	<.001
Pop-up menu	0.056	0.032-0.080	<.001
Navigational depth×Navigational lists per page	−0.031	−0.067 to 0.003	.08
Reading level	−0.050	−0.073 to −0.026	<.001
Large font	−0.036	−0.059 to −0.012	.003
Read level×large font	0.046	0.012-0.080	.008
Words per page	−0.020	−0.044 to 0.004	.10
Navigational depth×words per page	−0.038	−0.073 to −0.004	.03
Screen length	−0.043	−0.066 to −0.020	<.001
Hyperlinks per page×screen length	0.037	0.003-0.070	.03
Navigation areas per page	−0.018	−0.035 to 0.000	.05
Diagnostic group			
SSD[a] (vs others)	−0.059	−0.104 to −0.013	.01
Constant	0.604	0.563-0.644	<.001
Model 2: Design dimensions, diagnostic groups, and computer understanding			
Dimensions			
Navigational depth	−0.318	−0.348 to −0.289	<.001
Hyperlinks per page	−0.084	−0.108 to −0.060	<.001
Navigational lists per page	0.056	0.031-0.080	<.001
Navigational depth×navigational lists per page	−0.030	−0.065 to 0.004	.09
Reading level	−0.049	−0.073 to −0.025	<.001
Large font	−0.036	−0.059 to −0.013	.002
Read level×large font	0.048	0.0125-0.082	.005
Words per page	−0.021	−0.045 to 0.003	.09
Navigational depth×words per page	−0.038	−0.072 to −0.004	.03
Screen length	−0.044	−0.067 to −0.020	<.001
Hyperlinks per page×screen length	0.038	0.005 to 0.071	.03
Navigation areas per page	−0.017	−0.035 to 0.000	.05
Diagnostic group			
SSD (vs others)	−0.039	−0.110 to 0.032	.28
Self-reported level of computer understanding			
None/minimal (vs high)	−0.260	−0.328 to −0.192	<.001
Some (vs high)	−0.066	−0.112 to −0.019	.01
None/minimal understanding×navigational depth	0.089	0.467 to 0.130	<.001
Some understanding×navigational depth	0.024	−0.016 to 0.064	.24
Constant	0.687	0.640 to 0.735	<.001

[a]SSD: schizophrenia spectrum disorder.

Design Variables That Influence Time to Find Information: All Participants

Mixed regression models were created to identify the variables with main effects and interaction effects on the time it took participants to correctly find information. One of the models included the 12 design dimensions and 3 diagnostic groups (Table 5, model 1). Five of the main effects were the same variables as for the ability to find information (Table 4, model 1), and 4 of these at the higher level also had a negative effect on the time it took to find information. These variables had a negative effect on the time to find information: higher navigational depth, higher number of hyperlinks on a page, higher reading level, and pages longer than 1 screen length. The fifth variable, larger font, positively influenced time to find information, that is, reduced the amount of time needed. In addition, the following 3 variables had a negative effect at their higher levels: higher number of words per page (≥200), presence of a tool bar, and more words per hyperlink (≥6). In addition, SSD (vs others) was the only diagnostic group that entered the model, and it had a negative effect.

A second mixed regression model was developed (Table 5, model 2), which, in addition to the above variables, included the self-rated measure of participants' computer understanding. This entered the model as a significant variable. All of the other variables and 2-way interactions from model 1 remained in the model, including SSD.

Table 5. Multivariable models of the time to correctly find information.

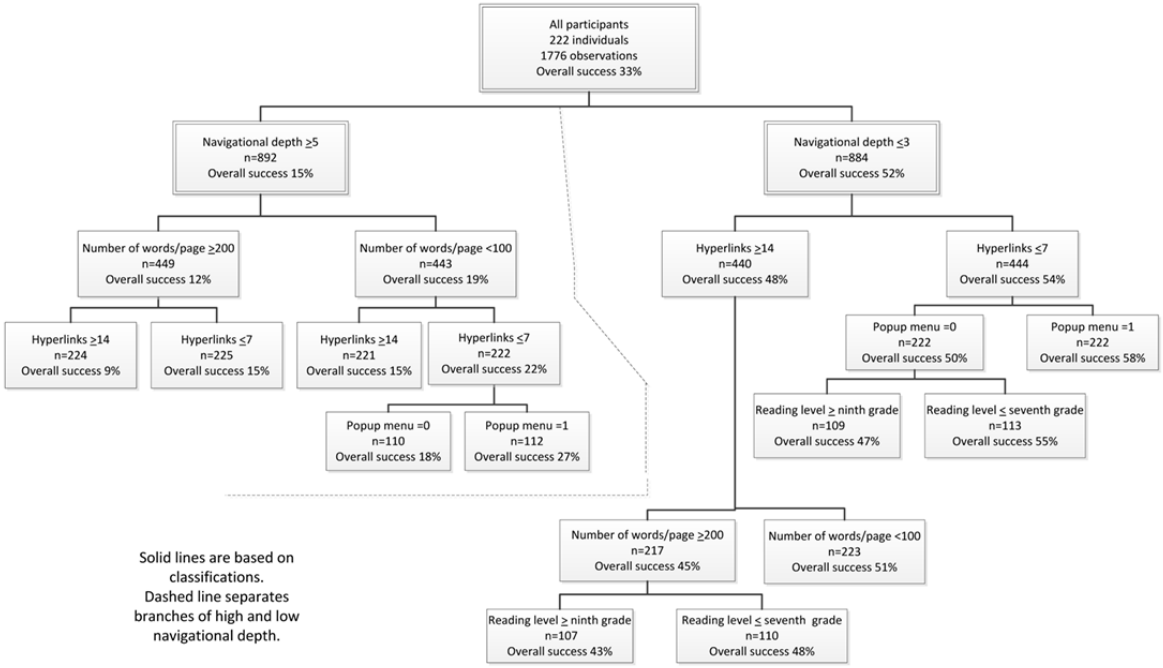
Variables	Coefficient	95% CI	Significance (*P* value)
Model 1: Design dimensions and diagnostic groups			
Dimensions			
Navigational depth	0.418	0.358 to 0.479	<.001
Hyperlinks per page	0.300	0.205 to 0.396	<.001
Reading level	0.208	0.131 to 0.285	<.001
Reading level×hyperlinks per page	−0.147	−0.263 to −0.029	.01
Navigational lists per page	0.057	−0.019 to 0.134	.14
Words per page	0.092	0.017 to 0.167	.02
Words per page×hyperlinks per page	−0.108	−0.219 to 0.003	.06
Toolbar	0.146	0.052 to 0.240	.002
Navigational lists per page ×toolbar	−0.138	−0.247 to −0.027	.01
Large font	−0.070	−0.122 to −0.017	.009
Screen length	0.054	0.002 to 0.106	.04
Words per hyperlink	0.092	0.013 to 0.170	.02
Words per hyperlink×toolbar	−0.101	−0.208 to 0.005	.06
Diagnostic group			
SSD[a] (vs others)	0.347	0.221 to 0.473	<.001
Constant	3.216	3.090 to 3.342	<.001
Model 2: Design dimensions, diagnostic groups, and computer understanding			
Dimensions			
Navigational depth	0.411	0.371 to 0.492	<.001
Hyperlinks per page	0.303	0.208 to 0.398	<.001
Reading level	0.200	0.123 to 0.276	<.001
Hyperlinks per page ×read level	−0.147	−0.264 to −0.031	.01
Words per page	0.096	0.021 to 0.171	.01
Hyperlinks per page×words per page	−0.112	−0.022 to −0.001	.048
Navigational lists per page	0.052	−0.025 to 0.129	.18
Toolbar	0.150	0.055 to 0.245	.002
Navigational lists per page×toolbar	−0.141	−0.250 to −0.030	.01
Large font	−0.073	−0.126 to 0.061	.006
Screen length	0.130	0.050 to 0.210	.002
Words per hyperlink	0.097	0.019 to 0.175	.02
Words per hyperlink×toolbar	−0.104	−0.211 to 0.003	.06
Diagnostic group			
SSD (vs others)	0.256	0.140 to 0.372	<.001
Self-reported level of computer understanding			
None/minimal (vs high)	0.555	0.399 to 0.712	<.001
Some (vs high)	0.257	0.116 to 0.397	<.001
Screen length×self-reported level of computer understanding			
None/minimal	−0.103	−0.242 to 0.034	.14
Some	−0.135	−0.255 to −0.016	.03
Constant	2.939	2.804 to 3.074	<.001

[a]SSD: schizophrenia spectrum disorder.

Designs That Increased or Reduced Usability in Terms of Ability to Find Information: CART Data Analysis Using All Participants and All Websites

The variable with the greatest influence on participants' abilities to successfully find information was navigational depth (Figure 2). For designs with deep navigation (ie, a navigational depth of ≥5 levels or screens), the average percent success was only 15%. For this subset, if in addition they had ≥200 words per page, and ≥14 hyperlinks per page the success rate decreased to 9%. This was the worst performing combination of the design elements tested in this category. The performance of designs with high navigational depth could be improved to a still exceedingly poor 27% success rate by using <100 words per page, ≤7 hyperlinks per page, and a pop-up menu to present the hyperlinks on a page.

Figure 2. Classification and regression trees analyses of all participants.



For websites designed with shallow navigation (ie, navigational depth ≤3 levels), the average success rate over all such designs and all participants was 52% (Figure 2). In addition, if they had ≤7 hyperlinks per page (54%) and used a pop-up menu to present the hyperlinks, the average success rate increased to 58%.

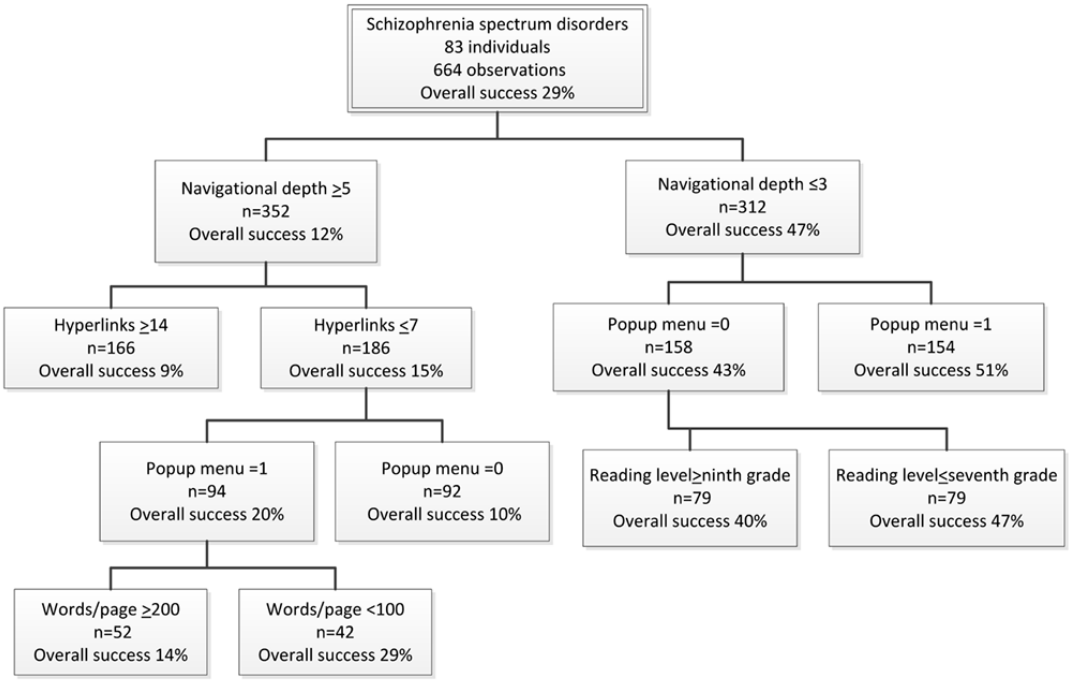
Of the designs with shallow navigational depth, the worst performing design combination had ≥14 hyperlinks per page (48%), ≥200 words per page (45%), and text with a reading level that was greater than or equal to ninth grade (43%).

Participants With SSD

Which Designs Make Usability Worse?

For individuals with SSD (n=83), the average success rate for all tested websites was relatively low (29%; Figure 3). The average success rate for all designs with deep navigational depth was only 12%. For these designs, if they had ≥14 hyperlinks per page, it fell to 9%. This was the worst performing design for this group of participants. Furthermore, other designs were apparently worse, but there were not enough subjects for the differences to be statistically significant.

Figure 3. Classification and regression trees analyses participants with schizophrenia spectrum disorder.



The highest performing designs of those with deeper navigational depth had a success rate of only 29%. This improvement was achieved by having ≤7 hyperlinks per page (15%), using a pop-up menu to present the hyperlinks on each page (20%), and having <100 words per page. Using these design elements in an overall poor design helped to improve performance but only modestly, and these designs were still quite poor.

Which Designs Make Usability Better?

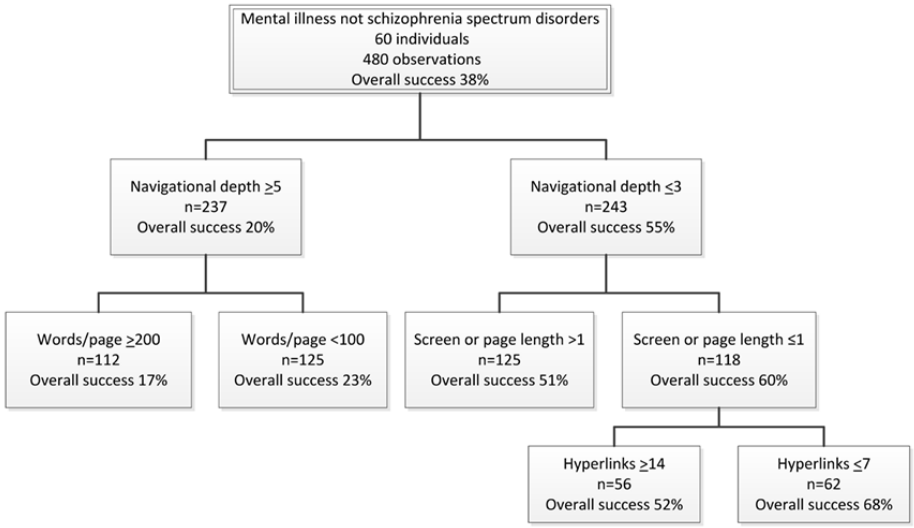
The average success rate for designs with shallow navigational depth was 47%. The best design within this set, with a success rate of 51%, was achieved by using a pop-up menu to present the hyperlinks. For websites with shallow navigation, the worst performing designs had a success rate of 40%. This occurred for websites with no pop-up menu and greater than or equal to ninth-grade reading level.

Participants With a Mental Illness Other Than SSD

Which Designs Make Usability Worse?

The average success rate for all tested designs with individuals with a mental illness other than SSD was 38% (Figure 4), and for all designs with deeper navigational depth, it was 20%. Performance was reduced to 17% for these latter designs if they also had ≥200 words per page. If the deeper navigational depth designs instead had <100 words per page, the success rate increased to 23%, which was the best performing deep navigational depth design for this group.

Figure 4. Classification and regression trees analyses of participants with a mental illness other than schizophrenia spectrum disorder.



Which Designs Make Usability Better?

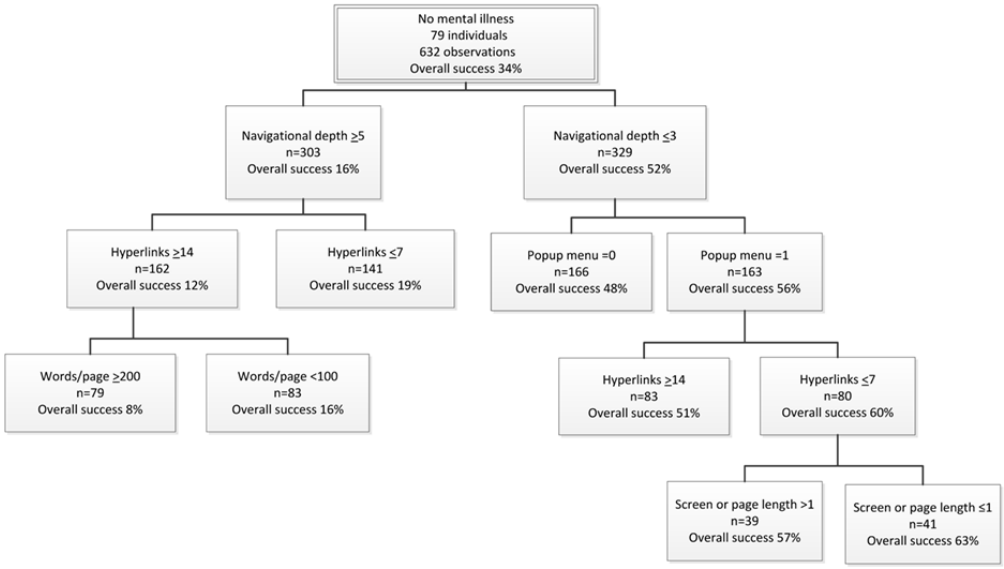
For all shallow navigational depth designs, the average success rate was 55%. Performance increased to 68% in designs that had page length ≤1 screen long (60%), that is, did not require scrolling to see all of the contents, and that had ≤7 hyperlinks per page. For websites with shallow navigation, the worst performing designs had a success rate of 51%. This occurred for designs where the page length was greater than one screen in length.

Participants With No Mental Illness

Which Designs Make Usability Worse?

For participants with no mental illness, the average success rate for all tested designs was 34% (Figure 5). For the deep navigational depth designs, the average success rate dropped to 16%. The performance dropped further to a low of 8% for websites having ≥14 hyperlinks per page (12%) and ≥200 words per page. Within the set of websites that had a deep navigational hierarchy, the best performing were those with ≤7 hyperlinks per page, which achieved an average success rate of only 19%.

Figure 5. Classification and regression trees analyses of participants with no mental illness.



Which Designs Make Usability Better?

The websites with shallow navigational depth (ie, ≤3 levels) had an average success rate of 52%. This is more than 3 (3.25) times the average success rate of designs with deep navigation (≥5 levels). The performance increased to 63%, if websites used a pop-up menu to present the hyperlinks (56%), had ≤7 hyperlinks per page (60%), and displayed information on ≤1 screen length (ie, no scrolling was used). The worst performing designs with a shallow navigational depth had a success rate of 48%. This occurred for websites with no pop-up menu.

Best and Worst Design Variables

Several design variables were commonly found in the best performing designs (Table 6) for each of the 4 participant groups (ie, SSD, mental illnesses other than SSD, no mental illness, and all participants); however, only shallow navigational depth was present in all 4 of these highest performing website designs. Design elements that always had a positive effect on usability were shallow navigational depth, presenting hyperlinks via a pop-up menu, ≤7 hyperlinks per page, and presenting the contents using a page length ≤1 screen long.

There were also several design elements that were commonly found on the worst performing website for each of the 4 participant groups (Table 6). High navigational depth was the only variable present in all 4 of these worst performing designs. The design elements that when present always had a detrimental effect on usability were high navigational depth, high words per page (≥200), ≥14 hyperlinks per page, page length >1 screen long, and high reading level (≥ninth grade).

Table 6. Best and worst performing designs and design variables for finding information.

Participant group	Design elements
The best performing websites had these design elements	
Diagnostic group	Dimensions (overall success rate for the design)
SSD[a]	Low navigational depth (≤3 levels) and pop-up menu (79/154, 51%)
Mental illness other than SSD	Low navigational depth (≤3 levels), page length ≤1 screen long, and ≤7 hyperlinks per page (42/62, 68%)
No mental illness	Low navigational depth (≤3 levels), pop-up menu, hyperlinks ≤7 per page, and page length ≤1 screen long (26/41, 63%)
All participants	Low navigational depth (≤3 levels), ≤7 hyperlinks per page, and pop-up menu (129/222, 58%)
The worst performing websites had these design elements	
SSD	High navigational depth (≥5 levels) and ≥14 hyperlinks per page (15/166, 9%)
Mental illness other than SSD	High navigational depth (≥5 levels) and ≥200 words per page (19/112, 17%)
No mental illness	High navigational depth (≥5 levels), ≥14 hyperlinks, and ≥200 words per page (6/79, 8%)
All participants	High navigational depth (≥5 levels), ≥200 words per page, and ≥14 hyperlinks (20/224, 9%)
The worst performing websites with low navigational depth (≤3 levels) had these design elements	
SSD	No pop-up menu and high reading level greater than or equal to ninth grade (32/79, 40%)
Mental illness other than SSD	Page length >1 screen long (64/125, 51%)
No mental illness	No pop-up menu (80/166, 48%)
All participants	≥14 hyperlinks per page, ≥200 words per page, and high reading level greater than or equal to ninth grade (46/107, 43%)
The best performing websites with high navigational depth (≥5 levels) had these design elements	
SSD	≤7 hyperlinks per page, pop-up menu and <100 words per page (12/42, 29%)
Mental illness other than SSD	<100 words per page (29/125, 23%)
No mental illness	≤7 hyperlinks per page (27/141, 19%)
All participants	<100 words per page, ≤7 hyperlinks per page, and pop-up menu (30/112, 27%)
Variables that, when present, always had a positive effect on performance	
Variable	Variable was also present in the best performing design for n of the 4 groups (the 3 diagnostic groups and all participants)
Low navigational depth	4
Pop-up menu used	3
≤7 hyperlinks	3
Page length ≤1 screen long	2
Variables that, when present, always had a negative effect on performance	
Variable	Variable was also present in the worst performing design for n of the 4 groups (the 3 diagnostic groups and all participants)
High navigational depth (≥5 levels)	4
≥200 words per page	2
≥14 hyperlinks per page	2
Page length >1 screen long	0 (none)
High reading level (≥ninth grade)	0 (none)

[a]SSD: schizophrenia spectrum disorder.

Discussion

Principal Findings

Overall, one of the key findings is that by varying the designs of eHealth applications in highly definable ways, it is possible to improve the effectiveness of an application for users, in terms of their success at finding information and time to find the desired information. The best designs, when compared with the worst, made a difference in the success of using a website, from 38% to 55% in the 3 diagnostic groups.

These data identify the depth of navigation, that is, the number of screens one needs to navigate through to reach the desired contents, as the most important variable for usability. This has also been identified as important by others [11,12,18,20]. When all participants' data were pooled, the average success rate for designs with a shallow hierarchy was 3.4 times higher than the average for those with a deep hierarchy. The worst performing design with a shallow hierarchy was still 2.5 times more successful than the best design with deep navigation. This principle was held across each of the 3 diagnostic groups as well. Although we found that the design environment (ie, the screen's overall design) influenced the effects of several design variables on usability, this was not the case for navigational depth. Its influence was invariable. The presence of other design elements influences how strong this effect is, but shallow is always better than deep, all else being equal. This indicates that a shallow navigational structure is an essential design feature for usability; therefore, creating a shallow navigational structure is an essential design challenge to address.

It should be noted that all participants were using these websites for the first time; however, cognitive limitations can restrict one's ability to comprehend complex designs and create a mental model of an application, even after repeated use [12]. Without an accurate mental model, it is harder to become a savvy user. Consequently, it is possible that the differences that were found between the participant groups were minimal differences in what might be found if the comparisons were made after repeated website use.

Our previous research found that it was more effective to have users scroll down a screen to obtain additional information than to navigate to a new screen (ie, *scrolling* was superior to *paging*) [17,20]. This study indicates that designs that used scrolling were never the highest performing designs and tended to be inferior to those that used paging. A recent design study found that users state a clear preference for paging over scrolling [30]. The findings of this study are consistent with this preference. As we have pointed out previously, scrolling is a poor design choice, but our work indicates that, in some circumstances, it may have certain advantages to paging, if paging significantly increases the depth of navigation. Depth is the most important single variable for usability, and this finding holds for all groups. Although users prefer paging, it is more convenient, and makes it less likely that content will be missed by users, adding pages is potentially perilous to performance, particularly for less savvy users who may be more likely to become confused or lost in a deeper hierarchy. The design lesson may be that the best designs, in terms of quantitative effectiveness and user preferences, will minimize the need to scroll and will, at the same time, use a shallow navigational structure that relies on paging.

This issue is relevant to other design elements, for example, the number of hyperlinks on a page. The findings of this study were clear; the lower number of links to contents, compared with the higher number, contributed to a superior design. This study did not attempt to determine whether there might be an *optimal* number or range of navigation links on a page. The design simply compared 2 disparate levels to determine whether this might be an important design variable, and it was an important design variable. Although using a higher number of links was clearly inferior, the most important design element for users' success was the number of screens they needed to navigate through. Therefore, in any given design, there could be a trade-off between the number of links on a screen and the navigational depth. The evidence of this study as well as past studies [17] indicate that having more links on a page, if it can significantly reduce the number of pages a user must navigate through, could be expected to be a more effective design choice, all else being equal.

For this study, we created a single question to allow participants to self-assess how savvy they felt they were with technology. We have referred to this variable as *computer understanding*. The computer understanding variable accounted for the effect of SSD in the regression model of the ability to find information. This indicates that it is not necessarily the characteristics of the illness per se, but rather familiarity and skills with technology that were a key reason for poorer performance by this group. This may be caused by less use of associated technologies. It is possible that those with SSD who have significant cognitive challenges will have greater difficulty becoming savvy and/or will benefit from having designs that specifically accommodate their cognitive needs. This study and our prior work indicate that specific design features can be helpful to those with SSD, who may have special cognitive needs, while not reducing the performance of others. A clear implication of the findings from this study is that training with eHealth technologies should improve less savvy individuals' performance with technology. This has been observed in our prior eHealth intervention studies [27].

The findings of this study indicate that the design environment can influence the impact of a design element, at least to some extent. This indicates that there is not necessarily just one route to designing a highly usable page or screen. However, the highest performing pages did have key similarities, and there were still noticeable differences in the usability between alternative *good* designs, that is, no 2 alternative designs had the same performance. This latter point supports the premise that there are certain design principles that seem to be fundamental to creating high-performance applications.

For all participants, when the measure of computer understanding was included in the regression model for the time required to find information, SSD (vs the other 2 diagnostic groups) remained a significant variable in the model. This would suggest that there is something in addition to how savvy one may be with these technologies that influences the time it takes to find information. This was not observed in the regression

model describing the ability to find information, where SSD (vs others) dropped out of the model, indicating that the relevance of computer understanding was a factor that was common to all groups combined. Deficits in processing have been found by others [31], and our own data show (manuscript in preparation) that the processing speed of those with SSD was slower than that of the other groups in the study. This might contribute to increasing the time that it takes to find information but does not influence the ability to find information.

Limitations

There are several limitations that need to be considered when interpreting these findings. Participants had to be able to read at the fifth-grade level to enter the study. The design needs of those with lower or even much higher reading levels may be different. The sample may have been too small to detect anything but main effects or very large interactions between the variables. The experimental design was limited to an evaluation of only 12 variables. Other variables may also be, and likely are, important. As a segment of the participants had little or no familiarity with technology, all participants were taken through a brief training that showed them how to use a mouse and demonstrated all of the website navigational elements they would encounter. This likely improved the abilities of these users and their performance compared with similar individuals in the general public who would not have such training.

Clinical Implications

One of the keys to successfully engaging consumers with eHealth treatments and services, particularly consumers with SMI and special cognitive needs, is to provide intuitive navigational designs. An evidence base of what works for creating effective designs and what does not will facilitate designs that can improve access, individualization, and treatment engagement for consumers. This should allow for the creation of eHealth services that are more usable; engaging; and, therefore, effective.

Conclusions

Seven of the key variables that influence how effective eHealth intervention designs are for those with and without mental health disorders are navigational depth, number of hyperlinks per page, presence of a pop-up menu, reading level, page length, the number of words per page, and a participant's skills with the technologies.

Future Research

Future work would benefit from a larger sample to better understand how these variables might interact with each other. This could allow the identification of specific ways in which the design environment interacts with design variables and synergistic or antagonistic effects of variable combinations. Each variable was studied at 2 levels only, which is common in a factorial design. Additional research could narrow the range of what is optimal. For example, a navigational depth of ≤3 was far superior to ≥5 levels, but the difference between 1, 2, and 3 levels might be considerable. However, it was not determined by these data. In addition, examination of variables other than these 12 variables would be very useful. A missing piece of this type of quantitative design research is user preferences for alternative designs. Coupling performance with preferences further advances the understanding of what works best for whom. We found that our single self-rated computer understanding question was very effective at measuring how savvy users were with the technologies. We are preparing a manuscript that fully describes this instrument and our findings. In addition, we are preparing a manuscript of our findings about the influence of the neurocognitive functions we collected from each participant on the effectiveness of the various designs and importance of the different design variables.

Acknowledgments

This work was supported by grants from the Veterans Administration, Rehabilitation Research and Development Service (D6180R), and Health Services Research and Development Service (HX001443). The contents of this paper do not represent the views of the US Department of Veterans Affairs or the United States Government.

Conflicts of Interest

None declared.

References

1. Dixon LB, Dickerson F, Bellack AS, Bennett M, Dickinson D, Goldberg RW, Schizophrenia Patient Outcomes Research Team (PORT). The 2009 schizophrenia PORT psychosocial treatment recommendations and summary statements. Schizophr Bull 2010 Jan 02;36(1):48-70 [FREE Full text] [doi: 10.1093/schbul/sbp115] [Medline: 19955389]
2. Kessler RC, Berglund PA, Bruce ML, Koch JR, Laska EM, Leaf PJ, et al. The prevalence and correlates of untreated serious mental illness. Health Serv Res 2001 Dec;36(6 Pt 1):987-1007 [FREE Full text] [Medline: 11775672]
3. Kreyenbuhl J, Slade EP, Medoff DR, Brown CH, Ehrenreich B, Afful J, et al. Time to discontinuation of first- and second-generation antipsychotic medications in the treatment of schizophrenia. Schizophr Res 2011 Sep;131(1-3):127-132 [FREE Full text] [doi: 10.1016/j.schres.2011.04.028] [Medline: 21576008]
4. Chien WT, Leung SF, Yeung FK, Wong WK. Current approaches to treatments for schizophrenia spectrum disorders, part II: psychosocial interventions and patient-focused perspectives in psychiatric care. NDT 2013 Sep:1463. [doi: 10.2147/ndt.s49263]

5. Kazdin AE. Addressing the treatment gap: a key challenge for extending evidence-based psychosocial interventions. Behav Res Ther 2017 Jan;88:7-18. [doi: 10.1016/j.brat.2016.06.004] [Medline: 28110678]
6. Higashi K, Medic G, Littlewood KJ, Diez T, Granström O, De Hert M. Medication adherence in schizophrenia: factors influencing adherence and consequences of nonadherence, a systematic literature review. Ther Adv Psychopharmacol 2013 Aug 04;3(4):200-218 [FREE Full text] [doi: 10.1177/2045125312474019] [Medline: 24167693]
7. O'Brien A, Fahmy R, Singh SP. Disengagement from mental health services. A literature review. Soc Psychiatry Psychiatr Epidemiol 2009 Jul 26;44(7):558-568. [doi: 10.1007/s00127-008-0476-0] [Medline: 19037573]
8. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016 Mar 28;12(1):157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
9. Griffiths K, Christensen H. Review of randomised controlled trials of internet interventions for mental disorders and related conditions. Clin Psychol 2006(1):29. [doi: 10.1080/13284200500378696]
10. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
11. Brunette MF, Ferron JC, Devitt T, Geiger P, Martin WM, Pratt S, et al. Do smoking cessation websites meet the needs of smokers with severe mental illnesses? Health Educ Res 2012 Apr;27(2):183-190 [FREE Full text] [doi: 10.1093/her/cyr092] [Medline: 21987478]
12. Rotondi AJ, Sinkule J, Haas GL, Spring MB, Litschge CM, Newhill CE, et al. Designing websites for persons with cognitive deficits: design and usability of a psychoeducational intervention for persons with severe mental illness. Psychol Serv 2007 Aug;4(3):202-204 [FREE Full text] [doi: 10.1037/1541-1559.4.3.202] [Medline: 26321884]
13. Bernard R, Sabariego C, Cieza A. Barriers and facilitation measures related to people with mental disorders when using the web: a systematic review. J Med Internet Res 2016 Jun 09;18(6):157 [FREE Full text] [doi: 10.2196/jmir.5442] [Medline: 27282115]
14. Friedman MG, Bryen DN. Web accessibility design recommendations for people with cognitive disabilities. TAD 2008 Jan 03;19(4):205-212. [doi: 10.3233/tad-2007-19406]
15. Ben-Zeev D, Kaiser SM, Brenner CJ, Begale M, Duffecy J, Mohr DC. Development and usability testing of FOCUS: a smartphone system for self-management of schizophrenia. Psychiatr Rehabil J 2013 Dec;36(4):289-296 [FREE Full text] [doi: 10.1037/prj0000019] [Medline: 24015913]
16. Bulger JR. A usability study of mental health websites with an emphasis on homepage design: performance and preferences of those with anxiety disorders. (Master's thesis) 2002 [FREE Full text]
17. Rotondi AJ, Eack SM, Hanusa BH, Spring MB, Haas GL. Critical design elements of e-health applications for users with severe mental illness: singular focus, simple architecture, prominent contents, explicit navigation, and inclusive hyperlinks. Schizophr Bull 2015 Mar 27;41(2):440-448 [FREE Full text] [doi: 10.1093/schbul/sbt194] [Medline: 24375458]
18. Ferron JC, Brunette MF, McHugo GJ, Devitt TS, Martin WM, Drake RE. Developing a quit smoking website that is usable by people with severe mental illnesses. Psychiatr Rehabil J 2011;35(2):111-116. [doi: 10.2975/35.2.2011.111.116] [Medline: 22020840]
19. Good A, Sambhanthan A. Accessing web based health care and resources for mental health: interface design considerations for people experiencing mental illness. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2014. [doi: 10.1007/978-3-319-07635-5_3]
20. Rotondi AJ, Spring M, Grady J, Simpson R, Luther J, Abebe KZ, et al. Use of a fractional factorial experiment to assess the e-healthcare application design needs of persons with dual diagnosis. J Dual Diagnosis 2012 Nov;8(4):277-282. [doi: 10.1080/15504263.2012.723454]
21. Black AC, Serowik KL, Schensul JJ, Bowen AM, Rosen MI. Build a better mouse: directly-observed issues in computer use for adults with SMI. Psychiatr Q 2013 Mar 19;84(1):81-92 [FREE Full text] [doi: 10.1007/s11126-012-9229-z] [Medline: 22711454]
22. Todd NJ, Jones SH, Lobban FA. What do service users with bipolar disorder want from a web-based self-management intervention? A qualitative focus group study. Clin Psychol Psychother 2013 Jun 20;20(6):531-543. [doi: 10.1002/cpp.1804] [Medline: 22715161]
23. Rotondi AJ, Spring MR, Hanusa BH, Eack SM, Haas GL. Designing eHealth applications to reduce cognitive effort for persons with severe mental illness: page complexity, navigation simplicity, and comprehensibility. JMIR Hum Factors 2017 Jan 05;4(1):1 [FREE Full text] [doi: 10.2196/humanfactors.6221] [Medline: 28057610]
24. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov 08;40(6):1244-1253 [FREE Full text] [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
25. Parameswaran R, Box GEP, Hunter WG, Hunter JS. Statistics for experimenters: an introduction to design, data analysis, and model building. J Market Res 1979 May;16(2):291. [doi: 10.2307/3150696]
26. First MB, Spitzer RL, Williams JBW, Gibbon M. Structured Clinical Interview for DSM-IV (SCID). American Psychiatric Association. 1997. URL: https://eprovide.mapi-trust.org/instruments/structured-clinical-interview-for-dsm-iv-axis-i-disorders [accessed 2012-02-12]

27. Rotondi AJ, Anderson CM, Haas GL, Eack SM, Spring MB, Ganguli R, et al. Web-based psychoeducational intervention for persons with schizophrenia and their supporters: one-year outcomes. Psychiatr Serv 2010 Nov;61(11):1099-1105. [doi: 10.1176/ps.2010.61.11.1099] [Medline: 21041348]
28. Kincaid JP, Fishburne RP, Rogers RL, Chissom BS, Station NA. Derivation of new readability formulas (automated readability index, Fog countFlesch reading ease formula) for navy enlisted personnel. Memphis, TN 1975. [doi: 10.21236/ada006655]
29. Gordon AD, Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Biometrics 1984 Sep;40(3):874. [doi: 10.2307/2530946]
30. Buitenweg DC, Bongers IL, van de Mheen D, van Oers HA, van Nieuwenhuizen C. Cocreative development of the QoL-ME: a visual and personalized quality of life assessment app for people with severe mental health problems. JMIR Ment Health 2019 Mar 28;6(3):12378 [FREE Full text] [doi: 10.2196/12378] [Medline: 30920381]
31. Fioravanti M, Bianchi V, Cinti ME. Cognitive deficits in schizophrenia: an updated metanalysis of the scientific evidence. BMC Psychiatry 2012 Jun 20;12(1). [doi: 10.1186/1471-244x-12-64]

Abbreviations

CART: classification and regression trees
SMI: severe mental illness
SSD: schizophrenia spectrum disorder

Edited by G Eysenbach; submitted 04.08.20; peer-reviewed by T Beentjes, A Hudon; comments to author 09.11.20; revised version received 07.12.20; accepted 18.12.20; published 24.03.21.

Please cite as:
Rotondi AJ, Grady J, Hanusa BH, Haas GL, Spring MR, Abebe KZ, Luther J, Gurklis J
Key Variables for Effective eHealth Designs for Individuals With and Without Mental Health Disorders: 2^12-4 Fractional Factorial Experiment
J Med Internet Res 2021;23(3):e23137
URL: https://www.jmir.org/2021/3/e23137
doi:10.2196/23137
PMID:33759796

Original Paper

Design Guidelines of a Computer-Based Intervention for Computer Vision Syndrome: Focus Group Study and Real-World Deployment

Youjin Hwang[1], MA; Donghoon Shin[1], BA; Jinsu Eun[1], MA; Bongwon Suh[2], PhD; Joonhwan Lee[1], PhD

[1]Human Computer Interaction and Design Lab, Seoul National University, Seoul, Republic of Korea

[2]Seoul National University, Seoul, Republic of Korea

Corresponding Author:
Joonhwan Lee, PhD
Human Computer Interaction and Design Lab
Seoul National University
64-406, Gwanak ro 1, Gwanak gu
Seoul, 08826
Republic of Korea
Phone: 82 02 880 6450
Email: joonhwan@snu.ac.kr

Abstract

Background: Prolonged time of computer use increases the prevalence of ocular problems, including eye strain, tired eyes, irritation, redness, blurred vision, and double vision, which are collectively referred to as computer vision syndrome (CVS). Approximately 70% of computer users have vision-related problems. For these reasons, properly designed interventions for users with CVS are required. To design an effective screen intervention for preventing or improving CVS, we must understand the effective interfaces of computer-based interventions.

Objective: In this study, we aimed to explore the interface elements of computer-based interventions for CVS to set design guidelines based on the pros and cons of each interface element.

Methods: We conducted an iterative user study to achieve our research objective. First, we conducted a workshop to evaluate the overall interface elements that were included in previous systems for CVS (n=7). Through the workshop, participants evaluated existing interface elements. Based on the evaluation results, we eliminated the elements that negatively affect intervention outcomes. Second, we designed our prototype system LiquidEye that includes multiple interface options (n=11). Interface options included interface elements that were positively evaluated in the workshop study. Lastly, we deployed LiquidEye in the real world to see how the included elements affected the intervention outcomes. Participants used LiquidEye for 14 days, and during this period, we collected participants' daily logs (n=680). Additionally, we conducted prestudy and poststudy surveys, and poststudy interviews to explore how each interface element affects participation in the system.

Results: User data logs collected from the 14 days of deployment were analyzed with multiple regression analysis to explore the interface elements affecting user participation in the intervention (LiquidEye). Statistically significant elements were the instruction page of the eye resting strategy (P=.01), goal setting of the resting period (P=.009), compliment feedback after completing resting (P<.001), a mid-size popup window (P=.02), and CVS symptom-like effects (P=.004).

Conclusions: Based on the study results, we suggested design implications to consider when designing computer-based interventions for CVS. The sophisticated design of the customization interface can make it possible for users to use the system more interactively, which can result in higher engagement in managing eye conditions. There are important technical challenges that still need to be addressed, but given the fact that this study was able to clarify the various factors related to computer-based interventions, the findings are expected to contribute greatly to the research of various computer-based intervention designs in the future.

(J Med Internet Res 2021;23(3):e22099) doi:10.2196/22099

KEYWORDS

computer-based intervention; computer vision syndrome; system interface; deployment study

Introduction

Background

As computer technologies advance rapidly, an increasing number of people spend their time in front of computer screens and mobile phones. According to the current population survey, 89% of US households have a computer, which includes smartphones, and 81% have a broadband internet subscription [1]. There is also computer use increase in the workplace. It is estimated that more than 75% of all jobs involve computer use [2]. Before personal computers revolutionized the workplace, office work had involved a range of activities, including typing, filing, reading, and writing. Each activity was adequately varied in the requirements of posture and vision, posing a natural "break" from the previous activity.

However, the introduction of personal computers has combined these tasks to where most can be performed without moving from the desktop, thereby improving quality, production, and efficiency, but also increasing computer-related health issues [2].

Computer vision syndrome (CVS) is one of the typical computer-related health issues [3-6]. Approximately 70% of computer users have CVS-related problems. The American Optometric Association defines CVS as the combination of eye and vision problems associated with the use of computers. The ocular complaints made by computer users typically include eye strain, eye fatigue, burning sensation, irritation, redness, blurred vision, and dry eyes, among others. The condition of a person experiencing one or more of these ocular complaints as a result of operating a computer and looking at a computer monitor is generally referred to as CVS. Symptoms of CVS also include extraocular symptoms, such as neck pain, back pain, and shoulder pain [7,8]. All these symptoms negatively affect the performance of everyday tasks, such as reading, driving, and computer use, which lowers quality of life [9].

Designing an appropriate intervention involving eye rest is one of the technology-based solutions to reduce the prevalence of CVS [10-13]. Among the various forms of interventions, a computer-based intervention is an appropriate form of intervention for CVS [9]. A computer-based intervention offers a great variety of options for assessing individuals, creating and delivering customized health messages, and providing individuals with the methods necessary to maintain or change their health-related behaviors [14]. Maximizing benefits and minimizing costs are important when designing health interventions, including digital health interventions such as computer-based interventions [15-19]. Inadequate design is one of the reasons for increasing costs in the process of modifying and re-evaluating interventions [19,20]. Therefore, a design study for effective computer-based interventions should be performed before introducing a prototype to users [21].

Objectives

In this study, we aimed to explore interface elements that affect participation in a computer-based intervention helping the eye resting behavior of users with CVS. For this, first, we investigated effective interface elements in existing computer-based interventions for users with CVS during a focus group study. Second, to further investigate the effectiveness of interface elements, we conducted a deployment study with our prototype LiquidEye having multiple interface options. Interface options included interface elements that were evaluated higher than the average in the focus group study. To demonstrate how the included elements affect user participation in eye resting behavior, we deployed LiquidEye in the real world with 12 participants.

Methods

Study Procedure

This study included a focus group study and a deployment study. To research existing computer-based interventions for vision protection, screening was conducted before the focus group. In this phase, researchers screened and listed the interface elements from existing systems. In the focus group consisting of an evaluation session and a redesign session, participants with CVS (n=7) evaluated each element by discussing its pros and cons. The evaluation session was conducted with a focus group interview. In the redesign session, participants discussed additional interface elements that could affect system participation and all participants evaluated the elements. With interface elements rated higher than the average, we developed LiquidEye and conducted a deployment study with 12 participants.

Screening Existing Systems

During the screening phase, researchers aimed to list feasible interface elements from existing systems. Our system selection criteria were designed through a four-step procedure. In step 1, we collected all previously studied systems regarding CVS in the human-computer interaction community or other related fields. There were systems such as EyeGuardian [22], EyePhone [23], DualBlink [9], LiDAR [5], BlinkBlink [24], and EyeProtector [25]. In step 2, apps from the app store or web-based interventions were collected, such as ProtectYourVision, RestOnTime, and EyeBreak. In step 3, systems with all the intervening effects were collected and preanalyzed by researchers. Finally, in step 4, three systems were selected to include as many elements as possible and minimize overlap between systems. Step 4 was conducted because showing too many systems in one place could confuse participants. In the end, three systems were included in the focus group session. One was a prior academic prototype (Eye Protector [25]) and two were obtained from a commercial app store (Protect Your Vision [26] and Rest on Time [27]).

Phase 1: Focus Group Study

To evaluate and discuss interface elements in existing computer-based interventions for CVS, we conducted a focus group discussion. We recruited seven participants (three male and four female participants; P1-P7) aged from 21 to 37 years (Table 1). All participants reported that they had frequently experienced CVS symptoms, such as blurred vision, dry eyes, eye strain, headache, neck pain, and back pain [4,7]. Recruited participants experienced at least three of these symptoms. Additionally, they were using a computer for at least 3 hours a

day. We recruited participants through the university's online community and clinical recruiting sites. Each participant was given a US $30 voucher after completing the final session. The focus group discussion consisted of two major sessions (evaluation session and redesign session).

Table 1. Information of the participants in the focus group study.

Participant number	Age (years)	Gender	Average computer use per day	Related symptoms
P1	29	Female	≥4 h	Blurred vision, dry eyes, eye irritation, and neck and back pain
P2	28	Male	≥4 h	Blurred vision, dry eyes, headache, and neck and back pain
P3	24	Female	≥4 h	Blurred vision, double vision, dry eyes, eye irritation, headache, and neck and back pain
P4	22	Female	≥2 h	Blurred vision, dry eyes, eye irritation, headache, and neck and back pain
P5	26	Male	≥3 h	Blurred vision, dry eyes, eye irritation, and neck and back pain
P6	21	Female	≥4 h	Blurred vision, dry eyes, and neck and back pain
P7	37	Male	≥4 h	Blurred vision, double vision, and dry eyes

Evaluation Session

In this session, participants were asked to evaluate the interface elements in existing systems with other participants. We introduced the three intervention systems for CVS. For each system, the included interface elements and their functions were described to participants in detail with a simulation. Thereafter, participants discussed each interface element in detail and mainly discussed its acceptability, which is an important consideration for health technologies and interventions [28]. Acceptable interventions make users more likely to engage and adhere to the system. Participants evaluated the acceptability of each interface element on a 7-point Likert scale.

Redesign Session

The objective of the redesign session was to explore additional interface elements that were not included in previous systems. Participants were instructed to draw their ideal intervention system on a paper. They were told that they could take some of the factors they evaluated in the previous session or add new ones if needed. This enabled us to further discover and evaluate new important elements that could not be considered in the previous session. The participants drew what they thought was a desirable system on a given blank sheet of paper. In this session, participants were also instructed to focus on the acceptability of the system. To consider as many factors as possible, a researcher did not give participants a preannounced time and waited until all participants had finished their drawings. It took a total of 20 minutes. Thereafter, the participants explained their desirable system and interface elements to other participants and two of the authors (YJ and DH). Newly suggested elements were listed by the authors, and participants evaluated these elements as they did in the evaluation session.

Phase 2: Deployment Study

LiquidEye: Computer-Based Intervention for Users With CVS

LiquidEye is a computer-based intervention system that helps achieve an adequate amount of eye rest among users with CVS. It helps users' eye resting behavior by providing an intervening screen with a black/white full-screen window. The goal of LiquidEye is to minimize vision-related symptoms and prevent the occurrence of vision-related symptoms. For this, LiquidEye attempts to manage a user's prolonged time of computer use, which is one of the critical causes of CVS [7].

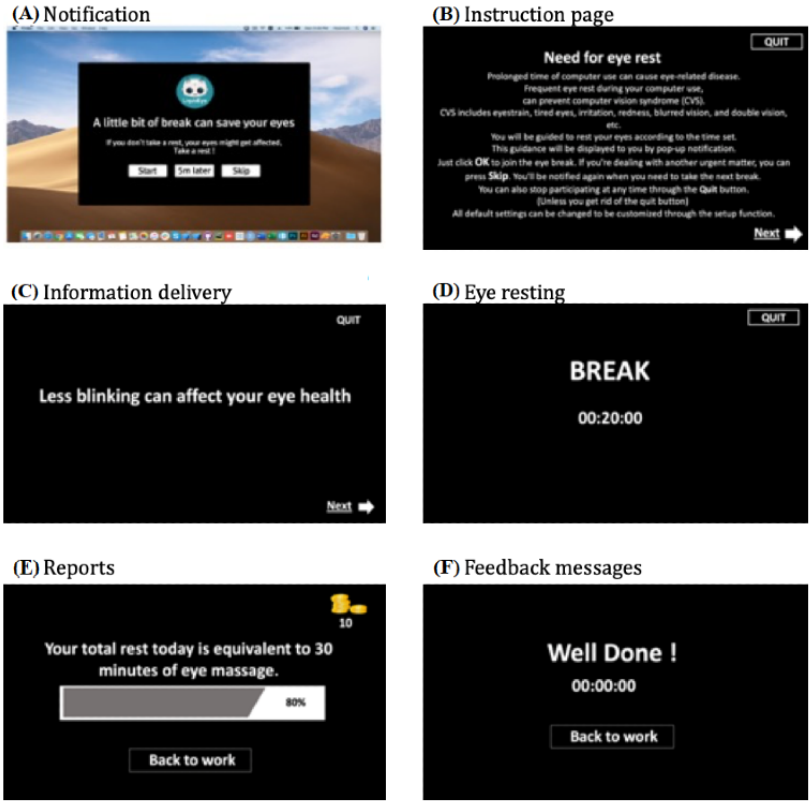
LiquidEye consists of multiple interface options. Interface options include interface elements that were evaluated higher than the average in the focus group study. As shown in Figure 1, users can select these options on the settings menu or can adjust the degree of each interface element (frequency, size, etc). Based on the user's settings, LiquidEye intervenes in prolonged computer use at the scheduled time with selected interface elements.

Figure 2 shows example scenarios of LiquidEye. At the scheduled time, a notification interface pops up and asks the user to participate in eye rest (Figure 2A). Depending on the user's customized settings, this element can accompany symptom-like visual effects. Users can choose among "start," "5 min later," and "skip." If the user clicks "start," LiquidEye records the user's behavior as "1 (participated)," and if the user clicks "skip," it records the user's behavior as "0 (did not participate)." When the user clicks "5 min later," the notification window interface pops up 5 minutes later. If the user turns off the notification window option, the eye resting scenario starts without a notification interface. Before starting eye rest, LiquidEye shows an instruction page that explains the need for eye rest and how to use LiquidEye (Figure 2B) and provides health information related to CVS (eg, less eye blinking can cause CVS, foods with beta-carotene can help improve eye conditions, etc) (Figure 2C). The instruction page and health information are optional depending on the user's customized settings. During eye rest, the word "break" appears and the word "look away" appears next and remains on the screen (Figure 2D). There is a timer in the middle that shows the remaining time for the user's eye resting behavior. Characters could be presented on the screen depending on the option settings. After the resting time, the user can receive a sound-based alarm

(optional). If the user quits the LiquidEye window (button on the top right) before the eye resting time is over, it records the user's behavior as "0 (did not participate)." If the user finishes eye resting without quitting, LiquidEye shows the user's eye resting accomplishment report for the day (Figure 2E). A rotated feedback message (compliment) can be sent to the user depending on the settings (Figure 2F).

Figure 1. Settings menu of LiquidEye. In the settings menu, users can customize their interface options for LiquidEye.

Figure 2. Example scenarios of LiquidEye. The interface elements (A-F) in the example scenarios are customizable in the settings menu.



LiquidEye was implemented on the MacOS system, and the system was developed with Swift in Xcode 10.3 IDE. After the development of LiquidEye, we conducted a 14-day real-world study with 12 recruited participants.

Participants and Procedure

A total of 12 participants (seven male and five female participants; A1-A12) aged from 22 to 40 years with CVS were recruited (Table 2). They all had at least three CVS-related symptoms and were using the computer for at least 3 hours a day. Participants were recruited through an online community. Participants in the focus group study did not overlap with participants in the deployment study. LiquidEye was used by participants for 2 weeks. Before participants started to use LiquidEye, they visited our lab to participate in a prestudy interview. Additionally, we helped participants to download LiquidEye on their personal computers and check if LiquidEye works properly on the device. To make participants use as many elements as possible, we instructed participants to use the system by changing the setting environment at least twice a day for an everyday task. Moreover, each time they changed the settings, they were asked to enter feedback for the previously used interface elements on an automatically appearing feedback page. Participants also visited the authors after the 14 days of participation for a postuse interview.

Table 2. Information of the participants in the 14-day deployment study.

Participant number	Age (years)	Gender	Average computer use per day	Number of reported CVS[a] symptoms	Occupation
A1	28	Female	≥4 h	4	Student
A2	31	Male	≥7 h	4	Data scientist
A3	22	Female	≥5 h	4	Student
A4	24	Female	≥3 h	3	Student
A5	23	Male	≥3 h	3	Office worker
A6	28	Female	≥5 h	3	Student
A7	28	Male	≥3 h	3	Data scientist
A8	24	Male	≥4 h	4	Student
A9	30	Male	≥7 h	3	Programmer
A10	33	Male	≥5 h	3	Office worker
A11	24	Female	≥4 h	3	Student
A12	40	Male	≥3 h	4	Office worker

[a]CVS: computer vision syndrome.

Thematic Analysis of Qualitative Data

Qualitative data that were collected through semistructured interviews and users' real-time feedback during LiquidEye use were integrated and analyzed. With these data, we conducted a thematic analysis to increase understanding. To build up themes from the data, we referred to the seven-element constructs of the Theoretical Framework for Acceptability developed by Sekhon et al [28]. These constructs consist of (1) affective attitude (how an individual feels about the intervention), (2) burden (perceived amount of effort that is required to participate in the intervention), (3) ethicality (extent to which the intervention has a good fit with an individual's value system), (4) intervention coherence (extent to which the participant understands the intervention and how it works), (5) opportunity cost (extent to which benefits, profits, or values must be given up to engage in the intervention), (6) perceived effectiveness (extent to which the intervention is perceived as likely to achieve its purpose), and (7) self-efficacy (participants' confidence that they can perform the behaviors required to participate in the intervention).

Statistical Analysis of Participation Depending on Interface Elements

With the user data logs collected during LiquidEye use, we conducted quantitative analysis. To investigate if each interface element significantly affects user participation in eye resting behavior suggested by the LiquidEye system (1: participated, 0: did not participate), multiple regression analysis was conducted. Interface elements were analyzed as independent variables, and participation was analyzed as a dependent variable. We conducted statistical analyses using R software (R Foundation for Statistical Computing).

Results

Results From Phase 1 (Focus Group Study)

Table 3 shows the list of interface elements evaluated in the focus group study and their ratings, including newly suggested interface elements in the redesign session during the focus group discussion. We classified these elements into subthemes and themes. During the classification process, we referred to the behavioral intervention technology model that involves frameworks integrating the conceptual framework into the technological framework [29].

Table 3. List and scores of interface elements that resulted from the focus group study.

Theme	Subthemes	Interface elements (score)[a]	Related systems
Behavior change strategies	Education	Instruction (4.0)	Rest on Time
Behavior change strategies	Goal setting	Goal selection (5.7)	Protect Your Vision
Behavior change strategies	Monitoring	Participation (4.8)	Rest on Time
Behavior change strategies	Feedback	Descriptive message (4.0) Comparative message (+) (4.2) Evaluative message (+) (4.2) Compliment message (4.8)	Rest on Time Protect Your Vision
Behavior change strategies	Reward	Monetary reward (+) (5.0) Score reward (+) (4.2)	N/A[b]
Elements	Information delivery	Health information (+) (6.0)	N/A
Elements	Notification	Popup (4.8) Full screen (4.3)	Rest on Time
Characteristics	Medium	Physical signal (−) (2.2) Sound (4.0) Screen based (6.2)	Protect Your Vision Eye Protector Rest on Time
Characteristics	Complexity	Rotated message (+) (6.0)	N/A
Characteristics	Aesthetics	An agent with robot appearance (4.0) An agent with expert appearance (+) (5.2) Spot effect (−) (2.8) Flashing effect (−) (2.8) Blurred effect (4.8) Symptom-like effect (+) (5.7)	Protect Your Vision Eye Protector
Workflow	User defined	Customization (+) (6)	N/A
Workflow	Conditions	20-20-20 (4.8) 60-5 (4.0) Just-in-time (+) (4.2)	Protect Your Vision Eye Protector Rest on Time

[a]Interface elements with low effectiveness are marked with "−," and interface elements newly added during the focus group discussion are marked with "+."

[b]N/A: not applicable.

Interface Elements With Low Effectiveness

Elements with low effectiveness (score below 4.0) were not included in LiquidEye. Interface elements with low effectiveness were visual effects of spot and flashing, and physical signal popups at the scheduled resting time (marked with "−" in Table 3). The spot is a feature that involves a colored dot icon intended to minimize interruption depending on the user's condition. Participant P6 made the following statement:

> *I don't think it's going to be noticeable. It only takes up a small part of the screen.*

The flashing effect was also discussed as below effective for the same reason as the spot. For physical signals, such as blowing wind toward the user's eyes, it was discussed as effective for grabbing the user's attention, but most participants rated it with a low score owing to its annoying interruption.

Newly Added Interface Elements

Additional interface elements were discussed in the focus group discussion (marked as "+" in Table 3). Symptom-like effects were suggested by participant P6. The participant made the following statement:

> *If the visual effect in the screen-based intervention come up with the CVS symptom like effects such as blurred vision or black spot, it will increase susceptibility to CVS, thus increase participation.*

The reward element was suggested by participant P2, participant P4, and participant P5. Participant P4 made the following statement:

> *Like playing the game, the rewards of making virtual money or getting high scores will affect not only early acceptability but also motivation for long-term use.*

For health information, participant P1, participant P5, and participant P6 indicated the need for this element. Participant P5 made the following statement:

> *Medical center does not usually give detailed eye-resting instructions. If we can get health information through this system, we can eventually make more efforts to improve CVS related symptoms.*

Participant P2 and participant P6 suggested a character with an expert-like appearance in the stage of eye resting instructions. Participant P6 made the following statement:

> *Expert-like character will increase the credibility of the information follows.*

A just-in-time function was suggested from the paper prototype of participant P4. Participant P4 made the following statement:

> *It would be more acceptable if the system has the function of avoiding important time such as meeting time.*

A customization option was suggested by most of the participants. Participant P1, participant P2, participant P4, participant P5, and participant P6 added customization options to their paper prototypes. Participant P6 made the following statement:

> *Different people have different demands for designs and functions, so it would be better if we could select the elements at the beginning of the system use.*

Rotation of messages in the system was suggested by participant P1, participant P2, participant P3, participant P6, and participant P7. Participant P1 made the following statement:

> *Rotated messages will make the system more useful.*

Additional Comments

After listing all interface elements in the focus group discussion, additional comments were collected to design LiquidEye. We conducted a focus group interview to discuss how the final elements (score above 4.0; to be implemented in LiquidEye) should be customized for the users. There existed several comments about varying frequency, varying interface size or design, and adding customization options. We present these results in line with the themes in Table 3.

For the education element (instruction on the system and how to use it), participants anticipated that the presence of this element matters more than how the element itself is organized. Some participants said they do not need it at all, while others said they want it to be for a specific period of time. Thus, two options, one with and one without the element, were implemented as customizable in LiquidEye.

For goal setting, which is setting resting frequency and the time of the day, most participants insisted that it should be customizable. In our case, reducing symptoms of CVS was our major clinical aim. To prevent CVS caused by prolonged computer use, clinical optometrists suggest users follow the 20/20/20 rule [30], which is that one should look at something 20 feet away for at least 20 seconds after 20 minutes of computer use [31]. However, since it is not easy to follow these guidelines, participants mentioned that they need flexibility with eye resting frequency and time, depending on their context. Therefore, we added an adjustable goal-setting element in our system.

Monitoring of participation was required or not depending on the individual. Thus, two options, one with and one without this element, were implemented as customizable in LiquidEye.

With regard to feedback, the kinds of feedback were not distinguishable. However, there was a difference between compliment feedback and others (descriptive message, comparative message, and evaluative message) according to most of the participants. Thus, we separated these two large categories in the setting options (Figure 1) and rotated the descriptive message, comparative message, and evaluative message.

There were opinions that there was no need to adjust the reward element depending on the context. If this element shows up in the system, it needs to keep showing up. All participants agreed that this element does not need to be customizable. Thus, it was kept as a basic setting.

The information delivery element (delivering health information) was required or not depending on the individual's preference. Thus, two options, one with and one without this element, were implemented as customizable in LiquidEye.

For the notification window, the size of the window can influence acceptability. Preference regarding the notification element (popup and full screen) varied. Participants commented as follows:

> *If my previous work environment is paused by the system anyway, I rather prefer full-screen.* [Participant #P1 and participant #P6]
>
> *Interruption has to be as small as possible.* [Participant #P7]

On the settings page of the system (Figure 1), the following four options were provided: full-screen notification window, mid-size window, small message popup on the top right of the screen, and none.

Physical signals, such as blowing winds, were eliminated from our final list since they were rated below our borderline (score 4.0). In the end, only the sound element was implemented in LiquidEye. Users can select the sound option or not in the settings menu. Regarding message rotation, all participants insisted that it is a necessary function for all time points. Therefore, this was set as a basic function. For the aesthetic element (presence of the character, color, etc), preferences varied among the participants. For this reason, we made it customizable for users. Users can select the color of the screen and the kind of character they like or can eliminate it.

Statistical Analysis With User Data Logs From Phase 2 (Real-World Deployment)

With LiquidEye, which was developed based on the focus group results, we collected user data logs during the 14 days of the experiment (n=680). To investigate interface elements that greatly affected the participation rate in the deployment, a multiple regression analysis was conducted with the users' overall data logs. Each interface element (total 14 elements) was analyzed as an independent variable, and participation (1: participated, 0: did not participate) was analyzed as a dependent variable. Table 4 shows the results from the multiple regression analysis. We present results in line with our themes and subthemes defined above. The relevant elements included the instruction page of the eye resting strategy, goal setting for eye

resting, compliment feedback after completing eye resting, mid-size popup window, and symptom-like visual effects that provide an alarm for the eye resting time.

Table 4. Results of multiple regression analysis.

Interface element (themes)	Interface element (subthemes)	Estimate	Standard deviation	*Z* value	*P* value
Intercept		−1.1114	0.6428	−1.73	.08
Education	Introduction page	0.6072	0.2469	2.46	.01
Goal setting	Default setting	−0.6647	0.2550	−2.61	.009
Goal setting	Adjusted setting	−0.0677	0.3730	−0.18	.86
Monitoring	Participation report	−0.1785	0.3697	−0.48	.63
Feedback	Default message	0.2460	0.2857	0.86	.39
Feedback	Compliment after eye resting	1.2977	0.3443	3.77	<.001
Information delivery	Health information	−0.6490	0.2824	−2.30	.02
Notification	Large-size window	−0.2572	0.3571	−0.72	.47
Notification	Mid-size window	−0.8873	0.3766	−2.36	.02
Notification	Small-size window	−0.6731	0.3798	−1.77	.08
Medium	Sound	0.2580	0.2297	1.12	.26
Aesthetic	Expert agent	−0.1277	0.3529	−0.36	.72
Aesthetic	Robot agent	−0.1270	0.3437	−0.37	.71
Aesthetic	Symptom-like effects with a notification window	0.7817	0.2714	2.88	.004

Discussion

Overview

Through two studies (ie, focus group study and deployment study), we explored the interface elements of computer-based interventions for CVS. Additionally, we collected real-world user data by deploying LiquidEye with customizable interface elements. With results from the deployment study, we could analyze how interface elements included in LiquidEye affected user participation with eye resting behavior. We will discuss the results while suggesting design guidelines for computer-based interventions for CVS.

Guidelines for Important Interface Elements

A summary of design guidelines for interface elements is presented in Table 5.

Based on our results, we will discuss the effect of each interface element on user participation with LiquidEye. We will also share the user feedback from the 14-day experiment with LiquidEye to discuss the results. We will first discuss the interface elements that greatly affected participation in the eye resting behavior, including the instruction page of the eye resting strategy, goal setting for eye resting, compliment feedback after completing eye resting, mid-size popup window, and symptom-like visual effects that provide an alarm for the eye resting time.

Regarding the instruction page, most participants agreed that it helped a lot at the beginning of the experiment, but was no longer needed after participants got used to it. As participants mentioned, the adaptation level affects the consequences of the interface element *instruction page* by increasing user intervention coherence or increasing user burden. System designers should consider how fast users adapt to the system and, at the same time, how easy or hard the system has been designed since these factors influence a user's need for the instruction page.

Table 5. Summary of design guidelines for interface elements.

Interface element (theme)	Example of interface element (subtheme)	Summary of design guidelines
Education	Introduction page	System designers should consider how fast users adapt to the system and, at the same time, how easy or hard the system was designed since these factors influence the user's need for an instruction page.
Goal setting	Default setting, adjusted setting (customizable)	System designers should consider the user's willingness to manage the eye condition since it decides a need for customization of goal settings. Default setting is the predefined setting regardless of the user's autonomy.
Monitoring	Participation report	This element can be a double-edged sword for the motivation of the user. It can increase or decrease the self-efficacy of the user depending on the level of participation.
Feedback	Default message, compliment after eye resting	Depending on the context of the user, it can be either effective or ineffective. However, the preference for this element was high among users.
Information delivery	Health information	System designers should consider the user's intention to manage the symptoms. If the user intention is high, the need for health information is also high at most times. However, low user intention can make users feel that this element is a burden.
Notification	Size of the window	The size of the popup influenced the forcefulness of the computer-based intervention. The full-screen notification with the high forcefulness was evaluated as most effective, but, at the same time, a high burden. Mid-size notifications positively affected user participation among other options.
Medium	Sound	The social context largely affected the user experience. Most of the participants insisted that it does not need to be in the system.
Aesthetic	Presence of characters (expert agent or robot agent) or visual effects (symptom-like effects)	Most of the time aesthetic elements rarely affect user participation, except when they strengthen the intervention effects by accompanying other intervention elements, such as the notification window in our case.

Goal setting for eye resting is another element that greatly affects user participation in eye resting behavior. It was interesting that the default setting for goal setting was relevant, while the adjusted setting was not. The default setting is a predefined setting based on the 20/20/20 rule (one should look at something 20 feet away for at least 20 seconds after 20 minutes of computer use) [30] for preventing or reducing CVS symptoms. Since the 20/20/20 rule is strict for long-time computer users as they have to rest three times per hour, we expected that customized settings (adjusted by users) would be more effective at increasing the participation rate. However, the customizable setting did not affect the user's participation rate according to our statistical data. From user feedback, we found out that customizable goal setting ("adjusted" in Table 4) can result in increased effectiveness or increased burden depending on the attitude of the user. Users evaluated the system with the goal setting element more effectively when they were willing to manage their symptoms compared with those who were not willing to manage their symptoms. For example, participant P7 with a low attitude level showed a negative opinion. This participant made the following statement:

> *It is too annoying to set goals since I feel no need to manage my symptoms.*

Compliment feedback after completing eye resting greatly affected user participation, but the qualitative results implied that it can sometimes be a burden for users. In particular, what users were doing right before the intervention affected the consequences of the feedback element. Participant A6 made the following statement:

> *I was working hard and then they told me to take a rest. I want to go back to my working environment as soon as the break is over. I don't feel like the extra things which are annoying and unnecessary.*

However, most participants said that this element plays a positive role when they are not busy. Participant A5 made the following statement:

> *Compliment feedback was really helpful. I always turn this element on as my basic setting. It makes me feel good!*

Regarding the interface element *notification window*, the mid-size popup window was related to user participation. The element was implemented in LiquidEye with four options (small-size popup, mid-size popup, full-screen popup, and no popup). Most of the participants insisted that the size of the popup influenced the forcefulness of LiquidEye. The full-screen popup with high forcefulness was evaluated as the most effective, but, at the same time, as having a high burden. Participant A10 made the following statement:

> *If I make up my mind to take a break anyway, I'd rather be forced to do it on time.*

On the other hand, participant A6 made the following statement:

> *Small pop-up is barely noticeable, which makes me miss the participation.*

Based on user feedback during and after the deployment study, we could infer that a mid-size popup window could be an alternative for the full-screen window and the small-size window with low effectiveness.

Another relevant interface element was symptom-like visual effects that provide an alarm for the eye resting time. An interesting opinion about this element was that its effectiveness

depends on the size of the window and how often the effect is being rotated. Participant A9 made the following statement:

> *When it comes to this element, how much it grabs my attention matters. When it accompanies a full-screen popup window, it does not grab additional attention, because the popup window already fills my whole screen. However, when it accompanies small or middle size popup window, it strengthens the system to grab additional attention.*

Guidelines on Other Interface Elements

We are going to discuss additional findings on other interface elements even though they were not found to be relevant. They did not show significance, but monitoring the user's participation and making a report on daily progress ("monitoring" in Table 4) can increase or decrease the self-efficacy of the user depending on the level of participation. Participant A3 made the following statement:

> *When I participated a lot, it was helpful for motivation but when I participated less, it was a burden to see.*

Additionally, participant A2 made the following statement:

> *I just want it to show me the number of times I participated, not the rate of participation. It only gets lower if I do not participate in 100 percent.*

The effectiveness of health information seems to depend on the intention to manage the symptoms. It could be effective if users are highly willing to manage their symptoms. Participant A11 made the following statement:

> *Getting this information makes me feel like I'm taking good care of my eyes. I spent more time thinking about my eyes.*

However, for those who have a low intention of participating in eye resting behavior, health information could be a bothersome interface element. Participant A2 expressed the following negative opinion:

> *Whether it is health information or anything else, a lot of text could be the burden to use the system.*

When it comes to a sound-based alarm ("sound" in Table 4), the social context largely affected the user experience. Most of the participants insisted that it does not need to be in the system. Participant A5 made the following statement:

> *I did not use it at least once since I always use my computer in my workplace.*

Few participants mentioned that it can be assistive but it must be optional.

For the character-like agent ("expert agent" and "robot agent" in Table 4), there rarely existed comments from users. Participant A1 made the following comment:

> *It is barely noticeable. It does not affect my participation.*

Additionally, users could customize their interfaces in the LiquidEye settings menu by themselves. This function of customization can increase effectiveness, but can be a burden depending on the clinical goal of the user. Most of the participants were satisfied with the customization options. Participant A4 made the following statement:

> *Depending on whether it is night or day, the desired setting is different since we are usually doing important things during the daytime and less important things during the nighttime.*

On the other hand, participant A7 made the following statement:

> *It is a burden to change the options frequently. I want it to just recommend me the best option which is not very disturbing.*

Avoiding work interruption was one of the major issues regarding user context. On the other hand, there is a need for a "right-on-time" intervention when it comes to clinical management of eye health. Participant A5 made the following statement:

> *I want it to show up right on time which is a most effective way for my eye health.*

However, most participants agreed with the idea that LiquidEye needs to avoid critical moments (eg, sharing the monitor with colleagues in the middle of a conference). Participant A4 made the following statement:

> *Adding the do-not-disturb function to the LiquidEye will make the system more acceptable.*

Application to Other Clinical Symptoms

For developing our computer-based intervention, CVS was chosen as our condition of interest. Before expanding our results to other clinical domains that require computer-based interventions, designers or system developers should consider the below-mentioned steps.

First, when choosing interface elements for computer-based interventions, the initial thing to do is feature the clinical aim and the usage aim [29,32]. System designers need to decide on these aims and the intervention medium before they choose the interface elements. Depending on the clinical aim and target behavior, the intervention medium can be different, which means that a computer-based intervention is not the best medium for all cases.

Second, understanding the target clinical group is crucial [33]. Even if the same interface element is being used, implementation strategies have to differ depending on users' unique features. Elements should be applied depending on the users' personal and health behavior–related factors, such as attitude, behavior intention, and ultimate health goals. If the target user group is too heterogeneous, a computer-based intervention can be an option since it offers a great variety of options for assessing individuals, creating and delivering customized health messages, and providing individuals with the methods necessary to maintain or change their health-related behaviors [14].

Third, the evaluation of a computer-based intervention has to be completed before final implementation in a large population. Even when two computer-based interventions use the same framework, the consequences can be different. In our study, we evaluated the interface elements in LiquidEye with statistical

analyses to better understand the consequences of choosing the interface elements.

If designers take all of the above points into consideration, our work is expected to decrease the cost of choosing interface elements in computer-based interventions by minimizing trial and error, even when implemented in other clinical domains.

Conclusions

To reduce the prevalence of CVS in computer users, designing appropriate interventions that induce eye rest is one of the technology-based solutions. In this study, we suggested design implications to consider when designing a computer-based intervention for CVS. The sophisticated design of a customizable interface can make it possible for users to use the system more interactively, which can result in higher engagement. Among the various interface elements that are being implemented in computer-based interventions for CVS, we found that the instruction page of the eye resting strategy, goal setting for eye resting, compliment feedback after completing eye resting, mid-size popup window, and symptom-like visual effects that provide an alarm for the eye resting time greatly affected user participation in the eye resting behavior. We manually defined how these elements affected user participation based on the framework of acceptability. In a further study, we will explore the opportunities of automated technologies, such as facial expression recognition [34], deep sentiment analysis [35], and gaze-tracking algorithms [36], to detect positive or negative user experiences with the computer-based intervention. There are important technical challenges that still need to be addressed, but given the fact that this study was able to clarify the various factors related to computer-based interventions, the findings are expected to contribute greatly to the research of various computer-based intervention designs in the future.

Acknowledgments

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A3A2075114).

Conflicts of Interest

None declared.

References

1. Ryan C, Lewis J. Computer and Internet Use in the United States: 2016. United States Census Bureau. 2018. URL: https://www.census.gov/content/dam/Census/library/publications/2018/acs/ACS-39.pdf [accessed 2021-03-13]
2. Neumann PG. Computer-Related Risks. Boston, MA, USA: Addison Wesley; 1994.
3. Widyantara P, Puspasari M. Breakpoint of Attention Media Evaluation as Countermeasure for Computer Vision Syndrome. In: Proceedings of the 2019 5th International Conference on Industrial and Business Engineering. 2019 Presented at: 5th International Conference on Industrial and Business Engineering; September 2019; Hong Kong p. 188-192. [doi: 10.1145/3364335.3364360]
4. Yan Z, Hu L, Chen H, Lu F. Computer Vision Syndrome: A widely spreading but largely unknown epidemic among computer users. Computers in Human Behavior 2008 Sep 01;24(5):2026-2042. [doi: 10.1016/j.chb.2007.09.004]
5. Comeau J, Godnig E. Computer use and vision. In: Proceedings of the 27th annual ACM SIGUCCS conference on User services: Mile high expectations. 1999 Presented at: 27th annual ACM SIGUCCS conference on User services: Mile high expectations; November 1999; Denver, CO, USA p. 31-34. [doi: 10.1145/337043.337078]
6. Sheedy JE. Vision problems at video display terminals: a survey of optometrists. J Am Optom Assoc 1992 Oct;63(10):687-692. [Medline: 1430742]
7. Blehm C, Vishnu S, Khattak A, Mitra S, Yee RW. Computer vision syndrome: a review. Surv Ophthalmol 2005;50(3):253-262. [doi: 10.1016/j.survophthal.2005.02.008] [Medline: 15850814]
8. Chawla A, Lim TC, Shikhare SN, Munk PL, Peh WCG. Computer Vision Syndrome: Darkness Under the Shadow of Light. Can Assoc Radiol J 2019 Feb;70(1):5-9. [doi: 10.1016/j.carj.2018.10.005] [Medline: 30691563]
9. Dementyev A, Holz C. DualBlink: A Wearable Device to Continuously Detect, Track, and Actuate Blinking For Alleviating Dry Eyes and Computer Vision Syndrome. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2017 Presented at: ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies; March 2017; USA. [doi: 10.1145/3053330]
10. Kawashima M, Uchino M, Kawazoe T, Kamiyashiki M, Sano K, Tsubota K. A field test of Web-based screening for dry eye disease to enhance awareness of eye problems among general Internet users: a latent strategy to promote health. J Med Internet Res 2013 Sep 27;15(9):e209 [FREE Full text] [doi: 10.2196/jmir.2198] [Medline: 24072379]
11. Inomata T, Nakamura M, Iwagami M, Midorikawa-Inomata A, Sung J, Fujimoto K, et al. Stratification of Individual Symptoms of Contact Lens-Associated Dry Eye Using the iPhone App DryEyeRhythm: Crowdsourced Cross-Sectional Study. J Med Internet Res 2020 Jun 26;22(6):e18996 [FREE Full text] [doi: 10.2196/18996] [Medline: 32589162]
12. Rodin A, Shachak A, Miller A, Akopyan V, Semenova N. Mobile Apps for Eye Care in Canada: An Analysis of the iTunes Store. JMIR Mhealth Uhealth 2017 Jun 14;5(6):e84 [FREE Full text] [doi: 10.2196/mhealth.7055] [Medline: 28615154]

13. Rono H, Bastawrous A, Macleod D, Bunywera C, Mamboleo R, Wanjala E, et al. Smartphone-Guided Algorithms for Use by Community Volunteers to Screen and Refer People With Eye Problems in Trans Nzoia County, Kenya: Development and Validation Study. JMIR Mhealth Uhealth 2020 Jun 19;8(6):e16345 [FREE Full text] [doi: 10.2196/16345] [Medline: 32558656]
14. Lustria MLA, Cortese J, Noar SM, Glueckauf RL. Computer-tailored health interventions delivered over the Web: review and analysis of key components. Patient Educ Couns 2009 Feb;74(2):156-173. [doi: 10.1016/j.pec.2008.08.023] [Medline: 18947966]
15. Ritterband LM, Andersson G, Christensen HM, Carlbring P, Cuijpers P. Directions for the International Society for Research on Internet Interventions (ISRII). J Med Internet Res 2006 Sep 29;8(3):e23 [FREE Full text] [doi: 10.2196/jmir.8.3.e23] [Medline: 17032639]
16. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006 Jun 23;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
17. Beintner I, Vollert B, Zarski A, Bolinski F, Musiat P, Görlich D, et al. Adherence Reporting in Randomized Controlled Trials Examining Manualized Multisession Online Interventions: Systematic Review of Practices and Proposal for Reporting Standards. J Med Internet Res 2019 Aug 15;21(8):e14181 [FREE Full text] [doi: 10.2196/14181] [Medline: 31414664]
18. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
19. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011 Jan 06;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
20. Etter J. Comparing the efficacy of two Internet-based, computer-tailored smoking cessation programs: a randomized trial. J Med Internet Res 2005 Mar 08;7(1):e2 [FREE Full text] [doi: 10.2196/jmir.7.1.e2] [Medline: 15829474]
21. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
22. Han S, Yang S, Kim J, Gerla M. EyeGuardian: a framework of eye tracking and blink detection for mobile device users. In: HotMobile '12: Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications. 2012 Presented at: Twelfth Workshop on Mobile Computing Systems & Applications; February 2012; San Diego, CA, USA p. 1-6. [doi: 10.1145/2162081.2162090]
23. Miluzzo E, Wang T, Campbell A. EyePhone: activating mobile phones with your eyes. In: Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds. 2010 Presented at: Second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds; August 2010; New Delhi, India p. 15-20. [doi: 10.1145/1851322.1851328]
24. Trutoiu LC, Carter EJ, Matthews I, Hodgins JK. Modeling and animating eye blinks. ACM Trans. Appl. Percept 2011 Aug;8(3):1-17. [doi: 10.1145/2010325.2010327]
25. Ho J, Pointner R, Shih H, Lin Y, Chen H, Tseng W, et al. EyeProtector: Encouraging a Healthy Viewing Distance when Using Smartphones. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services. 2015 Presented at: 17th International Conference on Human-Computer Interaction with Mobile Devices and Services; August 2015; Copenhagen, Denmark p. 77-85. [doi: 10.1145/2785830.2785836]
26. Protect Your Vision. Breaks For Eyes. URL: https://breaksforeyes.app [accessed 2021-03-13]
27. Rest on Time. Breaks For Eyes. URL: https://breaksforeyes.app [accessed 2021-03-13]
28. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
29. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
30. Anshel J. Visual Ergonomics Handbook. Boca Raton, FL, USA: CRC Press; 2005.
31. Reddy SC, Low CK, Lim YP, Low LL, Mardina F, Nursaleha MP. Computer vision syndrome: a study of knowledge and practices in university students. Nepal J Ophthalmol 2013;5(2):161-168. [doi: 10.3126/nepjoph.v5i2.8707] [Medline: 24172549]
32. Taylor TK, Webster-Stratton C, Feil EG, Broadbent B, Widdop CS, Severson HH. Computer-based intervention with coaching: an example using the Incredible Years program. Cogn Behav Ther 2008;37(4):233-246 [FREE Full text] [doi: 10.1080/16506070802364511] [Medline: 18803072]
33. Hetzroni OE, Tannous J. Effects of a Computer-Based Intervention Program on the Communicative Functions of Children with Autism. J Autism Dev Disord 2004 Apr;34(2):95-113. [doi: 10.1023/b:jadd.0000022602.40506.bf] [Medline: 15162930]

34. Leo M, Carcagnì P, Mazzeo PL, Spagnolo P, Cazzato D, Distante C. Analysis of Facial Information for Healthcare Applications: A Survey on Computer Vision-Based Approaches. Information 2020 Feb 26;11(3):128. [doi: 10.3390/info11030128]
35. Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: A survey. WIREs Data Mining Knowl Discov 2018 Mar 30;8(4):e1253. [doi: 10.1002/widm.1253]
36. Cazzato D, Leo M, Distante C, Voos H. When I Look into Your Eyes: A Survey on Computer Vision Contributions for Human Gaze Estimation and Tracking. Sensors (Basel) 2020 Jul 03;20(13):3739 [FREE Full text] [doi: 10.3390/s20133739] [Medline: 32635375]

Abbreviations

CVS: computer vision syndrome

Edited by R Kukafka; submitted 05.10.20; peer-reviewed by S Guness, M Leo; comments to author 18.11.20; revised version received 05.01.21; accepted 25.02.21; published 29.03.21.

Please cite as:
Hwang Y, Shin D, Eun J, Suh B, Lee J
Design Guidelines of a Computer-Based Intervention for Computer Vision Syndrome: Focus Group Study and Real-World Deployment
J Med Internet Res 2021;23(3):e22099
URL: https://www.jmir.org/2021/3/e22099
doi:10.2196/22099
PMID:33779568

Original Paper

A Uniquely Targeted, Mobile App-Based HIV Prevention Intervention for Young Transgender Women: Adaptation and Usability Study

Lisa M Kuhns[1,2], MPH, PhD; Jane Hereth[1,3], MSW, PhD; Robert Garofalo[1,2], MPH, MD; Marco Hidalgo[4], PhD; Amy K Johnson[1,2], PhD; Rebecca Schnall[5], MPH, RN, PhD; Sari L Reisner[6,7,8], ScD; Marvin Belzer[4], MD; Matthew J Mimiaga[6,8,9,10], MPH, ScD

[1]Ann & Robert H Lurie Children's Hospital of Chicago, Potocsnak Family Division of Adolescent and Young Adult Medicine, Chicago, IL, United States

[2]Department of Pediatrics, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Helen Bader School of Social Welfare, University of Wisconsin-Milwaukee, Milwaukee, WI, United States

[4]Division of Adolescent and Young Adult Medicine, Children's Hospital Los Angeles, University of Southern California, Los Angeles, CA, United States

[5]School of Nursing, Columbia University, New York, NY, United States

[6]Fenway Institute, Boston, MA, United States

[7]Division of General Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston, MA, United States

[8]Department of Epidemiology, Harvard T H Chan School of Public Health, Harvard University, Boston, MA, United States

[9]Center for Health Equity Research, Brown University, Providence, RI, United States

[10]Department of Psychiatry & Human Behavior, Alpert Medical School, Brown University, Providence, RI, United States

Corresponding Author:
Lisa M Kuhns, MPH, PhD
Ann & Robert H Lurie Children's Hospital of Chicago
Potocsnak Family Division of Adolescent and Young Adult Medicine
225 E Chicago Avenue, Box 161
Chicago, IL, 60611
United States
Phone: 1 312 227 7760
Email: lkuhns@luriechildrens.org

Abstract

Background: Young transgender women (YTW) are a key population for HIV-related risk reduction, yet very few interventions have been developed to meet their needs. Mobile health interventions with the potential for both efficacy and wide reach are a promising strategy to reduce HIV risk among YTW.

Objective: This study aims to adapt an efficacious group-based intervention to a mobile app, Project LifeSkills, to reduce HIV risk among YTW, and to test its acceptability and usability.

Methods: The group-based intervention was adapted to a mobile app, LifeSkills Mobile, with input from an expert advisory group and feedback from YTW collected during user-centered design sessions. A beta version of the app was then tested in a usability evaluation using a think-aloud protocol with debriefing interviews, recordings of screen activity, and assessments of usability via the Post-Study System Usability Questionnaire (PSSUQ) and the Health Information Technology Usability Evaluation Scale (Health-ITUES).

Results: YTW (n=8; age: mean 24 years, SD 3 years; racial or ethnic minority: 7/8, 88%) provided feedback on the app prototype in design sessions and then tested a beta version of the app in a usability trial (n=10; age: mean 24 years, SD 3 years; racial or ethnic minority: 8/10, 80%). Both usability ratings (Health-ITUES: mean 4.59, SD 0.86; scale range: 1-5) and ratings for satisfaction and accessibility (PSSUQ: mean 4.64, SD 0.90; scale range 1-5) were in the good to excellent range. No functional bugs were identified, and all mobile activities were deployed as expected. Participant feedback from the usability interviews indicated very good salience of the intervention content among the focal population. Participants' suggestions to further increase app engagement included adding animation, adding audio, and reducing the amount text.

Conclusions: We conclude that the LifeSkills Mobile app is a highly usable and engaging mobile app for HIV prevention among YTW.

(J Med Internet Res 2021;23(3):e21839) doi:10.2196/21839

KEYWORDS

transgender persons; HIV; mobile app; mHealth; mobile phone

Introduction

Background

HIV prevalence is disproportionately high among transgender women. Meta-analyses of previous studies indicate that the prevalence of laboratory-confirmed HIV infection among transgender women in the United States is between 19% and 28% [1,2], a level that is more than 30 times the odds of HIV-infection among the general population of adults of reproductive age. Younger and racial minority transgender women are particularly susceptible. A recent HIV testing initiative that involved over 1800 transgender women from 23 cities found the highest percentage of confirmed HIV seropositivity among Black transgender women (prevalence ratio: 3.13; 95% CI 1.45-6.78) [3]. In a previous study analyzing data from the local testing of over 500 transgender women (with no known previous positive HIV test results) in Miami, San Francisco, and Los Angeles, the highest number of cases was found among young transgender women (YTW) aged between 20 and 29 years (ie, 45%) [4].

Despite the high rates of HIV infection among transgender women, there are few interventions that have been developed specifically to address their underlying mechanisms of sexual risk [5]; only 2 interventions of sufficient scientific rigor and quality are included in the Center for Disease Control and Prevention's Compendium of HIV prevention interventions [6]. These two interventions include our LifeSkills group–based HIV prevention intervention for YTW aged between 16 and 29 years [7] and the Couples HIV Intervention Program, a face-to-face counseling intervention for adult transgender women and their partners [8]. Given the prevalence of HIV infection among YTW and the dearth of evidence-based interventions, there is a need to continue to develop and test HIV prevention interventions for YTW that are both efficacious and scalable.

Development and Testing of the LifeSkills Intervention

The LifeSkills intervention addresses the specific structural, developmental, and interpersonal challenges of HIV prevention among YTW [9,10]. It is based on empowerment theory [11-14] and was developed using a community-engaged approach to address the everyday social and structural drivers of risk among YTW [15-18]. The intervention promotes HIV-related information (education), motivation to reduce risk, and behavioral skills, consistent with the Information-Motivation-Behavioral skills model, one of the most parsimonious models of HIV-related behavior change [19-23], with the greatest utility for those at very high HIV risk [24].

The LifeSkills intervention was tested in a 2-city randomized controlled trial (RCT) and found to be efficacious with high rates of satisfaction. The RCT was conducted between 2012 and 2016 among 190 sexually active YTW [25]. The LifeSkills group had a 39.8% greater mean reduction in condomless sex acts (vaginal or anal) at 12-month follow-up compared with preventive care arm (0.71 vs 1.40; risk ratio: 0.60; 95% CI 0.50-0.72; $P<.001$) [7]. In addition, participants in the intervention arm reported high satisfaction with the curriculum: 98% of participants indicated they would refer a friend to receive LifeSkills [7]. However, some participants reported barriers to the uptake of the intervention, including problems attending group-based sessions because of competing priorities, difficult travel from other areas of the city or suburban locations, and scheduling conflicts. In addition, some participants expressed safety concerns regarding frequent travel to the study site, not because the site locations were unsafe but because travel on public transportation is often unsafe for YTW [26]. Furthermore, similar to group-based supportive therapy, a group-based HIV risk reduction intervention is not suitable for all who might benefit from it because of shyness, personality clashes, and the perception of lack of confidentiality.

Given these findings, we sought to extend the potential reach, accessibility, and privacy of the intervention via adaptation to a mobile app. Mobile health (mHealth) approaches (eg, health intervention via smartphones, tablet computers, or other mobile devices) for HIV prevention have the advantage of simple interface for users, accessibility anywhere internet access is available, and relative affordability and have been promoted specifically to reach stigmatized and disenfranchised populations [27,28]. A review of 62 HIV-specific mHealth studies found that most target HIV-positive adults (45%) and feature alerts and reminders (60%) as a primary health promotion strategy [27]. The findings suggest the potential for positive effects on health promotion across the HIV care continuum [27,28]. Although no HIV studies have specifically focused on HIV prevention among YTW, in a review of evidence for another key population, men who have sex with men, Schnall et al [29] found evidence that mHealth approaches, particularly web-based videos and education modules, show significant effects for both the reduction of HIV risk behavior and promotion of HIV testing. They also found that few mHealth approaches focus on the most recent biomedical prevention strategies, such as pre-exposure prophylaxis (PrEP), postexposure prophylaxis (PEP), and Treatment as Prevention (TasP). More specifically, evidence suggests that web-based interactive and educational approaches among youth aged between 13 and 29 years [30,31] are efficacious for delaying sexual initiation and increasing knowledge of HIV and sexually transmitted infections and condom self-efficacy [32]. In addition, gaming components of HIV-specific mHealth interventions are becoming increasingly

important, particularly among youth, because they are designed to be fun and may increase interaction and engagement, creating opportunities for learning [33].

This study aims to adapt the LifeSkills group–based intervention to a mobile platform and test its usability in the focal population, YTW aged between 16 and 29 years.

Methods

Overview

Extensive formative work to develop the LifeSkills intervention content was completed together with YTW [9]. Thus, our approach in this study was to adapt the original content to a mobile app. The adaptation of the LifeSkills Mobile app is based on best practices for the adaptation of evidence-based interventions and user-centered design [9,34]. Although focused primarily on new populations or settings, recommended adaptation strategies for evidence-based HIV prevention interventions include iterative processes, multiple methods and informants, and the preservation of core intervention components [35-37]. Thus, our adaptation process sought to preserve the components of the original intervention (ie, empowerment, information, motivation, and behavioral skill building), add biomedical prevention content (PrEP and TasP), and maintain its *look and feel* (eg, stories and scenarios of YTW) and unifying themes (ie, *House of LifeSkills*), while using an iterative, multimethod process with input from various key sources.

In this study, we adapted the in-person, group-based LifeSkills to mobile app and tested it in 3 phases: (1) prototype development, (2) design sessions and app development, and (3) usability testing (Multimedia Appendix 1).

The institutional review board at Ann & Robert H. Lurie Children's Hospital of Chicago approved all study procedures, and written consent was obtained from all study participants.

Prototype Development

In the first phase of the adaptation process, we translated the group-based intervention curriculum, which consisted of 6 sessions (4-5 activities each), into a paper-and-pencil app prototype with input from our group of 6 investigators and the study coordinator, who have extensive expertise in transgender health, youth-centered intervention development, HIV prevention and related behavioral intervention development efforts, and app development and informatics. In addition, 2 former facilitators of the group-based LifeSkills intervention served as community consultants to the project to inform the translation of the behavioral activation components (eg, activities, games, and scenarios) and *look and feel* to mobile app format. The development of the prototype occurred iteratively, with each draft reviewed in monthly meetings with the advisory group. The primary focus of prototype development was to maintain the core themes and activities of the original intervention, to draft interactive mobile activities to replace group-based activities, and to create engaging characters (to replace group facilitators). The end product was a paper-and-pencil prototype, which served as a blueprint of content for building the app.

Design Sessions and App Development

The draft paper-and-pencil prototype was provided to our software collaborator, One Cow Standing, as the first step in the app-building process. Wireframes were developed by One Cow Standing based on the prototype and the draft images of app characters and scenarios.

The draft prototype and images were then presented to the end users during the design sessions. Participants were identified and recruited from a registry of participants in previous studies. Individuals were eligible for design sessions based on (1) age between 16 and 29 years, (2) self-identity as transfeminine, (3) a history of condomless anal or vaginal sex, and (4) English speaking. We divided the prototype content into 3 subsections (of approximately equal length) and asked participants to review one-third of the content activity-by-activity, such that each subsection was reviewed by at least two participants. For each activity, we asked participants a set of semistructured questions focusing on user-centered aspects of adaptation to mobile, for example, whether they found the mobile app activities engaging (if so, why, and if not, why not), whether they understood the activity content (ie, wording, phrasing, and instructions), if they found it effective (ie, did it raise your awareness, motivate you, and introduce new skills), perceptions of the images and graphics, coverage of topics (ie, missed anything important), and any additional feedback or suggestions. Feedback from these sessions was recorded in extensive written notes for analysis. Data were rapidly analyzed for key actionable themes (eg, content to remove, revise, and add) [38], which were translated to modifications in the prototype.

We then implemented the app build in a series of design sprints over a 5-month period. Design sprints are time-constrained, phased periods in which specific programming work is completed, reviewed, and revised. In each sprint, programming of content was followed by testing of functional components and review of graphics and images by an abbreviated team of 4 alpha testers (LK, JH, and 2 community consultants), followed by discussion in weekly meetings with One Cow Standing. This process was facilitated by overlaying the test version of the app with BugHerd software (Splitrock Studio) [39], which was used to pinpoint functional bugs and areas in need of design improvements or modifications, with items placed in action queues to track completion. The final review was conducted by the entire advisory group of investigators in the last sprint with feedback and revision. The end product of this phase was a beta version of the app ready for usability testing.

Usability Testing

We performed usability testing on the beta version of the app to identify violations of usability principles and any potential obstacles to use. Eligibility for this phase of testing was the same as for design sessions; individuals who had participated in design sessions could also participate in usability testing, which provided an opportunity for feedback on the execution of design suggestions. Usability testing is an iterative process that involves testing the app and then using results to change the app to better meet users' needs. The usability testing phase was guided by the Health Information Technology Usability Evaluation Model (Health-ITUEM) to evaluate mHealth

technology [40], which includes a focus on quality, perceived usefulness, perceived ease of use, and user control.

We planned to execute at least two rounds of usability testing to identify potential problems and then correct and test the app, with rounds of testing added as needed to resolve all issues. Through a think-aloud protocol [41], each participant was asked to perform tasks that closely mirror the intended end use of the app. We used Morae software (TechSmith Corporation) [42] to record the interview audio and screenshots of the problem areas.

Participants completed a brief demographic and technology use survey and then rated the app's usability using validated instruments, including the Health Information Technology Usability Evaluation Scale (Health-ITUES) [43,44] and Post-Study System Usability Questionnaire (PSSUQ) [45]. The Health-ITUES evaluates tasks or expectations relevant to our usability dimensions, including impact, usefulness, ease of use, and user control. The PSSUQ measures accessibility and satisfaction. Both scales were rated on a 1- to 5-point scale, with higher scores indicating higher usability. We also measured satisfaction using the Client Satisfaction Questionnaire (CSQ) [46], which includes questions on overall quality and expectations. We measured acceptability with questions related to the use of language and the appeal of activities and artwork. These ratings were analyzed using measures of central tendency and dispersion.

Qualitative Data Collection and Analysis

At the end of the think-aloud session, we asked for additional open-ended feedback on perceptions of the app in a brief, semistructured interview. Interview data were analyzed using a thematic analysis approach [47,48] to identify key themes related to our usability dimensions, particularly perceptions of information quality, intention to use the app, satisfaction with the content, and potential impact of HIV risk reduction.

Results

Prototype Development

The original LifeSkills intervention consists of 6 group-based sessions (4-5 activities each), which were shortened and consolidated and then translated to a paper-and-pencil prototype consisting of the following 4 app-based modules: module 1: TransPride; module 2: Breaking Down Barriers; module 3: Educating Yourself, and module 4: Protecting Yourself. For the mobile app, we consolidated 2 modules in the original intervention (*communication and respect*, which introduces assertiveness skills training, and *skill building*, which applies these skills) into 1 mobile module, *Breaking Down Barriers* (module 2), which includes both training and application and eliminates some redundancy in the original intervention. Table 1 lists the general themes and intervention targets reflected in each module and the associated theoretical constructs. In addition, a unifying concept used in the original intervention, *The House of LifeSkills*, was maintained in the mobile version. This theme incorporates the language and symbols of the urban house and ball communities. The house and ball communities developed organically among largely gay, transgender, and gender-nonconforming youth of color in response to homelessness and family rejection, as a means of support as well as self-expression in the form of dance and other forms of performance art. The house and ball communities provide alternative family support in the face of the marginalization experienced by many urban transwomen [49,50]. In the LifeSkills intervention, the house and ball communities are a point of reference, as an example of the role of mentorship, community, and friendship within the transgender community. Symbols and language are used (sparingly) to contextualize the LifeSkills intervention as providing similar mentorship. All the primary themes, intervention targets, and theoretical constructs from the original LifeSkills intervention were maintained in LifeSkills Mobile.

Table 1. LifeSkills Mobile themes, intervention targets, and constructs.

Module	Theme	Intervention target	Theoretical constructs
1	TransPride	Self-concept, self-esteem, and sense of self in the context of societal marginalization	Empowerment and motivation
2	Breaking Down Barriers	Communication in the context of discrimination and violence	Skill building (effective communication and gaining access to resources)
3	Educating Yourself	Risks and benefits associated with sexual behavior and gender transition	Information (HIV and sexually transmitted infection risk and injection risk) and skill building (condom and needle use)
4	Protecting Yourself	Social context, experiences of abuse and violence, sexual risk, and substance use	Motivation (attitudes, norms, and intentions for safer sex) and behavioral skills (discussing sex and condoms with sexual partners and negotiating sexual safety, the context of drug and alcohol use, and sex work)

Design Sessions and App Development

Design Sessions

The prototype was reviewed in design sessions by 8 members of the target population of YTW (age: mean 24 years, SD 3 years; range 21-28 years; racial or ethnic minority: 7/8, 88%). The sessions lasted 75 minutes on average. User-specific themes, including engagement, understanding, effectiveness, perceptions of the images and graphics, and coverage of topics, are presented in detail below, with representative comments from participants.

With regard to engagement, participants were largely in agreement that the gaming approach with stickers and trophies

was fun, with one participant, a 24-year-old Black woman (D1), noting that they motivated her to keep going through the activities. In terms of understanding the content, participants described scenes and scenarios as realistic, relatable, and accurate. To improve understanding, participants suggested the use of *street* names for some of the illicit drugs in addition to the use of formal names (eg, *T* and *Tina* for methamphetamine and *leaf* and *weed* for marijuana). In terms of perceived effectiveness, participants commented that the salience and resonance of the content increased its effectiveness. For example, a 21-year-old woman who identified as gender nonconforming (D7) noted that the focus on content related to hormone use and other transition-related treatments would address common misconceptions. In addition, two participants, a 27-year-old multiracial woman (D5) and a 21-year-old White woman (D8), recommended that indicators of unhealthy relationships (in addition to the content on indicators of healthy relationships) be added because of the direct threat they represent to some transgender women (eg, abuse and physical violence) to increase its effectiveness. The perceptions of images and graphics were largely positive. Participants liked the images of characters and the artwork. For example, participants had mostly positive comments about the inclusiveness of gender expression depicted among characters (eg, early transition, gender variant, and feminine). In terms of coverage of topics, participants had positive comments about the addition of information on PrEP and PEP, noting that basic and accurate information on these topics is needed. In terms of additional comments and suggestions, one recurring theme was the perception that the app would improve the accessibility of the intervention to the target population of YTW. For example, a 25-year-old Black woman (D3) stated that the app "would be good for those who might not be comfortable coming to a group...Now I'm looking at a screen, I can go back to it...It's going to be dope."

A 21-year-old multiracial woman (D4) shared that the app would reach young women who did not know about the original LifeSkills groups or other resources. She stated:

> *Everyone is on their phone. So [the app] can reach people that didn't even know that any of this was going on.*

Overall, the findings from this phase suggested successful adaptation of the content to the mobile prototype with some suggestions for improved language and additions to content. The prototype was updated with these changes to inform the full development of the alpha and beta versions of the app.

Design Sprints

LifeSkills Mobile is a web app developed for mobile use (ie, not a native app intended for distribution through app stores). The decision to develop a web app was based on the higher cost of developing native apps and the additional complications posed by native apps, including distribution rules through app stores that change frequently. A web app is cross-platform and does not require platform-specific modifications, thereby increasing flexibility. Although browsers cannot use all the native features of a mobile device, they can present rich and engaging content and use device features, such as the camera (which is used in module 1). There are also no design elements in LifeSkills Mobile that require native app features. Finally, because it is not platform specific, it can reach a wider audience as all modern mobile devices have web browsers and are not limited to mobile phones for use; it can be accessed via a tablet or desktop computer.

The LifeSkills Mobile intervention was designed using JavaScript as the underlying programming language for both the client and server portions of the app. JavaScript is a cross-platform language that runs on all major operating systems. It is also the de facto programming language of the World Wide Web. This allows the code to be uniform and consistent as well as to easily make updates and modifications. The server component consists of a representational state transfer app program interface developed in Node.js (OpenJS Foundation) [51] using the Express framework. The database is MongoDB (MongoDB Inc) [52]. The web client interface was developed using the Vue.js (Vue.js) [53] framework.

Table 2 lists the final set of activities associated with each module in the beta version of the app. Module 1: TransPride sets the tone for the entire intervention and is devoted to empowerment themes. It includes transgender-specific success stories and a structured activity to create a participant's own aspirational story (Icon Promo) and finally a goal setting activity that is updated throughout the intervention. Module 2 focuses on effective communication to meet basic needs (Communication Styles) and applies it to a scenario of escalating violence (Assertiveness Under Pressure), an activity to deconstruct and activate basic aspects of communication (both verbal and body expression), and a second scenario in a clinical care setting in the context of transspecific discrimination (Health care Barriers to create awareness of situations, triggers, feelings, and actions). Similarly, module 3 includes activities to raise awareness of sexual, injection, and other risk and practice correct steps for condom and injection use. Module 3 also includes information and resources specific to PEP and PrEP as well as HIV transmission, HIV care, medication adherence, and TasP. In module 4, specific contexts of risk are addressed, including relationship contexts (eg, power dynamics), challenges to skill building (eg, disclosure, substance use, commercial sex work, and partner selection and negotiation), and ending with an action plan to activate goals set and reviewed throughout the intervention.

The intervention activities, which consisted of scenarios, educational material and games, and role plays, were adapted to a mobile format using automated responses, drop-down menus, comics, and gamification to engage participants. For example, the app allows the user to input their own goals via drop-down menus, which differ based on personal experience and circumstances; these goals are challenged throughout the intervention by scenarios that differ between individuals (eg, substance use and involvement in sex work). Gamification is the use of game-like rewards and incentives to increase motivation and sustain intervention engagement over time [54]. We used two elements of gamification: (1) educational games and (2) persuasive games [55]. Education games are small games or minigames that use elements such as drag-and-drop, speed, and memory. Simulation was used for role-playing, in which

participants had to envision themselves in different roles and choose the correct actions and dialogs to solve a task [56]. Persuasive games are used to motivate participants; participants have a tally bar (Multimedia Appendix 2) and supportive messages to motivate them to complete LifeSkills. Participation is incentivized with stickers for each activity and trophies for completion of modules. The YTW characters represent diverse gender presentations, racial or ethnic backgrounds, and HIV statuses. Similar to the group-based version of LifeSkills, these characters present scenarios to open each session and motivate participation in the next session.

Table 2. LifeSkills Mobile app modules and activities.

Module	Activity number	Activity title
Welcome	1	Welcome to LifeSkills Mobile
TransPride	2	Icon Promo (Pride)
TransPride	3	Why LifeSkills? (Risk and Protection)
TransPride	4	Get Your Life! (Setting Goals)
Breaking Down Barriers	5	Communication Styles
Breaking Down Barriers	6	Assertiveness Under Pressure
Breaking Down Barriers	7	Health care Barriers
Educating Yourself	8	Jeopar-T (Transition)
Educating Yourself	9	HIV True False
Educating Yourself	10	Bucket List! (Low, Medium, High Sex Risk)
Educating Yourself	11	Condom Use Steps
Educating Yourself	12	Safer Injection Steps
Protecting Yourself	13	Healthy Relationships
Protecting Yourself	14	Disclosure
Protecting Yourself	15	What are Your Boundaries
Protecting Yourself	16	In The Mix (Alcohol, Drugs)
Protecting Yourself	17	Sex Work and One Night Stands
Protecting Yourself	18	Setting Good Boundaries
Protecting Yourself	19	Red Flag Green Flag (Partner Negotiation and Selection)
Protecting Yourself	20	Action Plan (Accomplishing Goals)

Usability Testing

The beta version of the app was completed, and usability evaluation was conducted among end users (n=10; age: mean 24 years, SD 3 years; range 21-28 years; racial or ethnic minority: 8/10, 80%; ever traded sex: 6/10, 60%; ever homeless: 10/10, 100%; General Educational Development/high school diploma or less: 7/10, 70%). Of 10 participants, 7 completed the design sessions. Of 10 participants, 9 indicated they use the internet *almost constantly*, and the same number (9/10) participants reported using the internet via smartphone or handheld device. A total of 5 participants completed the first round of usability testing, followed by an additional 5 participants who completed the second round of testing.

Usability ratings were in the good to excellent range (Health-ITUES: mean 4.59, SD 0.86), including on subscales for impact, usefulness, ease of use, and user control. Satisfaction and accessibility on the PSSUQ was also strong (mean 4.63, SD 0.90; range 1-5). On the CSQ, half of the participants rated the overall quality of the app as *good* and half as *excellent*. In total, 9 of the 10 participants indicated that they would *definitely* refer a friend to use the app.

No functional bugs were identified in either round of testing, and all activities were deployed as expected. Small errors in grammar, language, or image placement were identified by participants in the first round of testing, which were corrected for redeployment in the next iteration of usability testing. In addition, although there were no functional bugs identified, 2 activities, activity number 10 (Bucket List!) and number 17 (Sex Work and One Night Stands), caused confusion during completion in round 1 of testing. For activity number 10, items describing sexual acts are categorized by participants as *low*, *medium*, or *high risk*; however, participants reported confusion about how to categorize some risk behaviors. Brief definitions of each risk category were added for clarification. For activity number 17, strategies to protect oneself in contexts of commercial sex work are matched to risk contexts; however, participants expressed difficulty making matches because of perceived overlap in strategies. For this activity, redundancy in strategies was eliminated, and language was edited for clarity. With these changes, the intervention was redeployed with subsequent participants in round 2 of testing, with no further reports of problems understanding and completing the activities.

Qualitative Findings

The analysis of substantive feedback and suggestions from participants in usability interviews formed 5 major categories related to (1) the *House* theme and characters, (2) the language and tone of the intervention, (3) the content and features, (4) the audience, and (5) how they might engage with the app. We review each of these themes in detail below, with representative comments from the participants.

House Theme and Characters

Most women responded positively to the House theme and look of the characters. In addition to making it more engaging and interesting, women also stated that they thought the characters looked realistic and relatable and that drew them in. A 26-year-old Black woman stated:

> *I like the characters. I like the words they've chosen to use. It's like, I feel like they're speaking out to me... I feel like it gets me. Yeah, I like the connection of this app.* [U4]

A 23-year-old Black woman (U10) stated about the housemother: "she reminded me of myself." Although most participants expressed positive feedback about the House theme, several noted that for YTW who were not familiar with the house and ball communities, it might be confusing.

Language and Tone

Women offered positive feedback regarding the language used. Many participants stated that they liked that the app used terms and phrases they used and that it had a casual, conversational tone. Regarding using slang terms for sexual acts, a 25-year-old Black woman stated:

> *I actually like that because it's giving me the real deal...It's kind of like we all talk, you know...So to actually see this on the app, it makes it more exciting. Like, "Oh, now it's talking my language.*" [U3]

A 29-year-old Black woman (U1) stated that the language "made me feel real. It made me feel like I could connect to it."

Regarding using language that is specific to the transgender community, a participant stated:

> *I love how they talk about the T a lot. "Girl, watch the T.*" [U4]

Overall, women shared that the tone and relatable language made them feel comfortable and affirmed. A participant stated:

> *I haven't seen an app that was dedicated to us that actually felt like was for us. This actually feels like it's for us and it actually felt like somebody is listening to us, they're hearing us, and they want to get our thoughts on how we can better ourselves. They want to help us better ourselves. That's what it feels like when I was doing the lessons in this app.* [U4]

Participants shared that they also appreciated that the app talked openly about sex work because it was realistic and would make people who were engaged in sex work feel more comfortable. A 21-year-old person who identifies as gender nonconforming stated:

> *I like the fact that you all have a client—like the sex workers and stuff.* [U7]

They went on to state:

> *I guess that making people feeling more comfortable about talking about stuff like that...You all know what's really going on and stuff like that.*

Regarding an activity related to drug use, participants shared that they thought the tone came across as nonjudgmental and informative. A participant stated:

> *I thought you were just like informative and educational. You didn't make it seem like it was bad to do drugs. It's okay to do them, but you informed us on the effect that it have on the body and stuff like that.* [U4]

Content and Features

In general, women shared that they liked the app because they learned something new. A participant said:

> *This is just one level that I'm on, and you've already hit me somewhere because I've already seen three things that I can apply to my everyday life.* [U4]

Women shared that the content was very important for transgender women. A multiracial 27-year-old woman stated:

> *I actually think you guys are touching a lot of topics that we don't, in the community ourselves, don't talk about sometimes... Just like discrimination period. I don't see a lot of young people having those conversations... But it's teaching you that you can overcome it and it's giving you more positivity and helping you to understand your life personally.* [U2]

Participants also shared that they liked the features and interactive nature of the app. Participant U2 stated:

> *It's actually really interactive and fun. I like it. I could see me playing it.*

Participant U4 said:

> *I really like everything about the app. It's interesting and I find it fun. It's like a game I'm playing, but educational. You can play a game, and you can learn at the same time. So, I like it.*

Many also shared that they felt motivated by earning stickers and trophies, and they liked how they looked.

Negative feedback about the content included that there was too much text on each page, suggesting that textual content should be edited or broken up into separate pages. They also suggested the inclusion of images of the characters or other images on every page to hold the user's interest. In addition, several participants suggested that having audio of the housemother's voice would be more engaging.

Audience

Women shared that they thought the app would be helpful for younger transgender women, including women who were still in high school or aged less than 18 years. Participant U9 said:

> *There are a lot of different things that would have been very, very good for me to know when I was younger that I did not that were in here.*

A 24-year-old Black woman (U6) stated that, to her, the app seemed too juvenile, but that it would be good for younger girls. U10 shared that she thought the app would be especially helpful for younger girls who are just transitioning and coming out and are more impressionable. She stated:

> *I really think it would have a bigger effect from 14 to age 18.*

App Engagement

Participants in this study only completed portions of the app; several stated that they would do it in chunks, whereas others said they might complete the entire app in one sitting. Participant U2 stated:

> *I'd do like, four or five [activities] at a time and come back later on, like in an hour and do four or five more.*

A 21-year-old multiracial woman (U5) said that she would only spend 30 minutes at a time but would return to it throughout the day. Participant U9 stated:

> *I'd probably go through the whole thing in one sitting, and then just want to be able to reference things in the future.*

Discussion

Principal Findings

In summary, in the usability testing of the LifeSkills Mobile app, we found no functional bugs, and all mobile activities deployed as expected. Participant feedback in the usability interviews indicated very good salience of the intervention content for the focal population.

To adapt the intervention, we used multiple methods and informants, including input from substantive experts, transgender women who had previous experience in the delivery of the in-person intervention, and members of the target population. This expertise was critical to reducing the length of the intervention for mobile deployment while also maintaining the core intervention targets, activities, and themes. The success of the development of the initial prototype, with fidelity to the original intervention components, was reflected in the largely positive comments from design session participants, who had relatively few critiques of its content and structure.

The process of building the app, characterized by intensive design sprints with iterative development and testing sprints by members of our advisory group and consultants, led to a bug-free beta version of the app, as demonstrated by the findings of the usability trial. Among testing by 10 YTW, we found no functional bugs and high usability ratings. Substantive input from YTW included improving the clarity of language and reducing the amount of redundancy in 2 activities, which resulted in improved performance.

The LifeSkills Mobile app is characterized by user-centered design features. The advantages of mobile delivery include standardized delivery of content, which increases fidelity, but also flexibility to tailor content to subgroups, giving control of some aspects of content to the user. These features allow more flexibility in the delivery of content than was possible in the in-person group-based intervention.

Taken together, qualitative and quantitative findings suggest that the LifeSkills Mobile app met the usability expectations of the Health-ITUEM [40] for quality, perceived usefulness, perceived ease of use, and user control. In general, feedback from participants in the usability interviews was largely positive and validated the app themes, intervention content, engagement strategies, and language and tone. Perhaps most importantly for engagement and uptake of the app, comments and feedback from semistructured interviews regarding the app activities and content, conducted as part of the usability trial, indicate resonance and salience among YTW. This indicates the success of the app in reflecting the real-life contexts and scenarios faced by YTW. It is important to note that the use of terms and their understanding, particularly slang terms, may differ by group (eg, age, race, and ethnicity) and may also change over time. Thus, it is important to monitor any problems with understanding and relevance and update the app as needed for future use.

In addition, participants reported positive impressions of the gamification approach, and the *House* theme of the app was generally well received. We recognized that this theme might not be applicable or relatable to all YTW. Although a unifying theme, as in the group-based LifeSkills intervention, it is used subtly and mostly symbolically as a tool to communicate the idea of positive mentoring; therefore, direct involvement in or even knowledge of the house and ball communities is generally not necessary to understand its meaning.

LifeSkills Mobile is among the first mHealth interventions that have been developed to meet the specific and unique needs of YTW to reduce sexual risk for HIV acquisition and transmission. It joins other such mHealth initiatives, focusing on transgender women, which are in the development or testing phases [57-60].

Limitations

The LifeSkills intervention content was developed with and for a largely ethnic minority YTW at high risk for HIV acquisition or transmission; thus, the LifeSkills Mobile app may not appeal to YTW at lower risk. In addition, the app may be particularly effective for younger transwomen, as noted by users. The reasons for this are that the app includes transition-related topics, including access to medical care and tips for safer hormone injections, which are especially important for those who are still in the process of gender transition. However, these tips are important at any age, even after transition, particularly in the context of HIV prevention because of the risk associated with needle use. YTW aged between 16 and 20 years were recruited for this study but were not included in our final sample. The lack of inclusion of YTW aged between 16 and 20 years limits our ability to generalize with confidence to this younger age group. The small sample size limits generalizability. However, the original intervention content, which forms the core of the mobile app, was developed and tested among transgender women aged between 16 and 29 years in 2 previous studies

[7,10]. Dynamic app content, including audio, video, and animation, have not yet been added to the app because of resource limitations but are planned for a future iteration. Similarly, the next version of the app will include the implementation of security features, including usernames and passwords, necessary for log-in and browser time-outs. We plan to test the efficacy of LifeSkills Mobile in a future RCT.

Conclusions

Overall, the adaptation and usability findings indicate that the LifeSkills Mobile app is a highly usable and engaging mobile app to address HIV prevention and related mechanisms of risk among YTW.

Acknowledgments

The authors thank Ebonii Warren-Watts and Kacey Draine for their contributions to the mobile app design. The authors also thank Nathanael Nash for assistance with data collection, Marbella Cervantes for data management, and Jenny Olson for study coordination assistance. This study was funded by the National Institute of Mental Health of the National Institutes of Health (NIH) under award number R56 MH113684. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The study sponsor had no role in the design of the study.

Authors' Contributions

LK and MM contributed to the design of all aspects of this study. LK and JH drafted the manuscript and tables. LK, JH, RG, MH, AJ, RS, SR, MB, and MM reviewed the manuscript. The manuscript has been read and approved by all of the authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Lifeskills mobile phases of app development.
[PNG File , 10 KB - jmir_v23i3e21839_app1.png]

Multimedia Appendix 2
Lifeskills mobile app map.
[PNG File , 293 KB - jmir_v23i3e21839_app2.png]

References

1. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of HIV in transgender women: a systematic review and meta-analysis. Lancet Infect Dis 2013 Mar;13(3):214-222. [doi: 10.1016/S1473-3099(12)70315-8] [Medline: 23260128]
2. Herbst JH, Jacobs ED, Finlayson TJ, McKleroy VS, Neumann MS, Crepaz N, et al. Estimating HIV prevalence and risk behaviors of transgender persons in the United States: a systematic review. AIDS Behav 2008 Jan;12(1):1-17. [doi: 10.1007/s10461-007-9299-3] [Medline: 17694429]
3. Pitasi MA, Clark HA, Chavez PR, DiNenno EA, Delaney KP. HIV Testing and Linkage to Care Among Transgender Women Who Have Sex with Men: 23 U.S. Cities. AIDS Behav 2020 Aug;24(8):2442-2450. [doi: 10.1007/s10461-020-02804-6] [Medline: 32020510]
4. Schulden JD, Song B, Barros A, Mares-DelGrasso A, Martin CW, Ramirez R, et al. Rapid HIV testing in transgender communities by community-based organizations in three cities. Public Health Rep 2008;123 Suppl 3:101-114 [FREE Full text] [doi: 10.1177/00333549081230S313] [Medline: 19166094]
5. Garofalo R, Kuhns LM, Reisner SL, Mimiaga MJ. Behavioral Interventions to Prevent HIV Transmission and Acquisition for Transgender Women: A Critical Review. J Acquir Immune Defic Syndr 2016 Aug 15;72 Suppl 3:S220-S225 [FREE Full text] [doi: 10.1097/QAI.0000000000001084] [Medline: 27429186]
6. Centers for Disease Control and Prevention. Compendium of evidence-based interventions and best practices for HIV prevention updated March 4, 2020. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/research/interventionresearch/compendium/index.html [accessed 2020-04-03]
7. Garofalo R, Kuhns LM, Reisner SL, Biello K, Mimiaga MJ. Efficacy of an Empowerment-Based, Group-Delivered HIV Prevention Intervention for Young Transgender Women. JAMA Pediatr 2018 Oct 01;172(10):916. [doi: 10.1001/jamapediatrics.2018.1799]
8. Operario D, Gamarel KE, Iwamoto M, Suzuki S, Suico S, Darbes L, et al. Couples-Focused Prevention Program to Reduce HIV Risk Among Transgender Women and Their Primary Male Partners: Feasibility and Promise of the Couples HIV Intervention Program. AIDS Behav 2017;21(8):2452-2463 [FREE Full text] [doi: 10.1007/s10461-016-1462-2] [Medline: 27334464]

9. Cotten C, Garofalo R. Project life skills: Developing the content of a multidimensional HIV-prevention curriculum for young transgender women aged 16 to 24. Journal of HIV/AIDS & Social Services 2016 Mar 22;15(1):99-113. [doi: 10.1080/15381501.2015.1074976]
10. Garofalo R, Johnson AK, Kuhns LM, Cotten C, Joseph H, Margolis A. Life skills: evaluation of a theory-driven behavioral HIV prevention intervention for young transgender women. J Urban Health 2012 Jun;89(3):419-431 [FREE Full text] [doi: 10.1007/s11524-011-9638-6] [Medline: 22223033]
11. Fawcett SB, Paine-Andrews A, Francisco VT, Schultz JA, Richter KP, Lewis RK, et al. Using empowerment theory in collaborative partnerships for community health and development. Am J Community Psychol 1995 Oct;23(5):677-697. [doi: 10.1007/BF02506987] [Medline: 8851345]
12. Glanz K, Rimer, BK, Viswanath K. Health behavior and health education: Theory, research, and practice. 4th ed. San Francisco: Jossey-Bass, Wiley Inc; 2008.
13. Rappaport J. Terms of empowerment/exemplars of prevention: toward a theory for community psychology. Am J Community Psychol 1987 Apr;15(2):121-148. [Medline: 3604997]
14. Shearer NBC. Relationships of contextual and relational factors to health empowerment in women. Res Theory Nurs Pract 2004;18(4):357-370. [doi: 10.1891/rtnp.18.4.357.64094] [Medline: 15776756]
15. Garofalo R, Deleon J, Osmer E, Doll M, Harper GW. Overlooked, misunderstood and at-risk: exploring the lives and HIV risk of ethnic minority male-to-female transgender youth. J Adolesc Health 2006 Mar;38(3):230-236. [doi: 10.1016/j.jadohealth.2005.03.023] [Medline: 16488820]
16. Garofalo R, Osmer E, Sullivan C, Doll M, Harper G. Environmental, Psychosocial, and Individual Correlates of HIV Risk in Ethnic Minority Male-to-Female Transgender Youth. Journal of HIV/AIDS Prevention in Children & Youth 2007 May 15;7(2):89-104. [doi: 10.1300/J499v07n02_06]
17. Wilson EC, Garofalo R, Harris DR, Belzer M. Sexual risk taking among transgender male-to-female youths with different partner types. Am J Public Health 2010 Aug;100(8):1500-1505 [FREE Full text] [doi: 10.2105/AJPH.2009.160051] [Medline: 20622176]
18. Wilson EC, Garofalo R, Harris RD, Herrick A, Martinez M, Martinez J, et al. Transgender female youth and sex work: HIV risk and a comparison of life factors related to engagement in sex work. AIDS Behav 2009 Oct;13(5):902-913 [FREE Full text] [doi: 10.1007/s10461-008-9508-8] [Medline: 19199022]
19. Carey MP, Maisto SA, Kalichman SC, Forsyth AD, Wright EM, Johnson BT. Enhancing motivation to reduce the risk of HIV infection for economically disadvantaged urban women. J Consult Clin Psychol 1997 Aug;65(4):531-541 [FREE Full text] [doi: 10.1037//0022-006x.65.4.531] [Medline: 9256553]
20. Fisher JD, Fisher WA, Williams SS, Malloy TE. Empirical tests of an information-motivation-behavioral skills model of AIDS-preventive behavior with gay men and heterosexual university students. Health Psychol 1994 May;13(3):238-250. [Medline: 8055859]
21. Fisher J, Fisher W. A. , Cornman, D. H.. Clinician-delivered intervention during routing clinical care reduces unprotected sexual behavior among HIV-infected patients. Journal of acquired immune deficiency syndromes (1999). .doi 10.1097/01.qai.0000192000.15777.5c 2006;41:44-52. [doi: 10.1097/01.qai.0000192000.15777.5c]
22. Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychological Bulletin 1992;111(3):455-474. [doi: 10.1037/0033-2909.111.3.455]
23. Fisher W, Fisher J. A general social psychological model for changing AIDS risk behavior. In: Pryor JB, Reeder GD, editors. The social psychology of HIV infection. Hillsdale, NJ: Erlbaum; 1993:127-153.
24. Kalichman SC, Picciano JF, Roffman RA. Motivation to reduce HIV risk behaviors in the context of the Information, Motivation and Behavioral Skills (IMB) model of HIV prevention. J Health Psychol 2008 Jul;13(5):680-689. [doi: 10.1177/1359105307082456] [Medline: 18519441]
25. Kuhns LM, Mimiaga MJ, Reisner SL, Biello K, Garofalo R. Project LifeSkills - a randomized controlled efficacy trial of a culturally tailored, empowerment-based, and group-delivered HIV prevention intervention for young transgender women: study protocol. BMC Public Health 2017 Sep 16;17(1):713. [doi: 10.1186/s12889-017-4734-5] [Medline: 28915919]
26. Hereth J, Kuhns L, Reisner S, Mimiaga M, Garofalo R. Lessons learned from a randomized controlled trial of a uniquely targeted group-based HIV prevention intervention among young transwomen. 2015 Presented at: Annual meeting of the American Public Health Association (APHA); 2015; Chicago, IL.
27. Catalani C, Philbrick W, Fraser H, Mechael P, Israelski DM. Open AIDS J 2013;7:17-41 [FREE Full text] [doi: 10.2174/1874613620130812003] [Medline: 24133558]
28. Devi BR, Syed-Abdul S, Kumar A, Iqbal U, Nguyen P, Li YJ, et al. mHealth: An updated systematic review with a focus on HIV/AIDS and tuberculosis long term management using mobile phones. Comput Methods Programs Biomed 2015 Nov;122(2):257-265. [doi: 10.1016/j.cmpb.2015.08.003] [Medline: 26304621]
29. Schnall R, Travers J, Rojas M, Carballo-Diéguez A. eHealth interventions for HIV prevention in high-risk men who have sex with men: a systematic review. J Med Internet Res 2014;16(5):e134 [FREE Full text] [doi: 10.2196/jmir.3393] [Medline: 24862459]

30. Bull S, Pratte K, Whitesell N, Rietmeijer C, McFarlane M. Effects of an Internet-based intervention for HIV prevention: the Youthnet trials. AIDS Behav 2009 Jun;13(3):474-487 [FREE Full text] [doi: 10.1007/s10461-008-9487-9] [Medline: 19037719]
31. Markham CM, Shegog R, Leonard AD, Bui TC, Paul ME. +CLICK: harnessing web-based training to reduce secondary transmission among HIV-positive youth. AIDS Care 2009 May;21(5):622-631 [FREE Full text] [doi: 10.1080/09540120802385637] [Medline: 19444671]
32. Guse K, Levine D, Martins S, Lira A, Gaarde J, Westmorland W, et al. Interventions using new digital media to improve adolescent sexual health: a systematic review. J Adolesc Health 2012 Dec;51(6):535-543. [doi: 10.1016/j.jadohealth.2012.03.014] [Medline: 23174462]
33. Hightow-Weidman LB, Muessig KE, Bauermeister JA, LeGrand S, Fiellin LE. The future of digital games for HIV prevention and care. Curr Opin HIV AIDS 2017 Sep;12(5):501-507. [doi: 10.1097/COH.0000000000000399] [Medline: 28692490]
34. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
35. Wingood GM, DiClemente RJ. The ADAPT-ITT model: a novel method of adapting evidence-based HIV Interventions. J Acquir Immune Defic Syndr 2008 Mar 01;47 Suppl 1:S40-S46. [doi: 10.1097/QAI.0b013e3181605df1] [Medline: 18301133]
36. Solomon J, Card JJ, Malow RM. Adapting efficacious interventions: advancing translational research in HIV prevention. Eval Health Prof 2006 Jun;29(2):162-194. [doi: 10.1177/0163278706287344] [Medline: 16645183]
37. McKleroy VS, Galbraith JS, Cummings B, Jones P, Harshbarger C, Collins C, ADAPT Team. Adapting evidence-based behavioral interventions for new settings and target populations. AIDS Educ Prev 2006 Aug;18(4 Suppl A):59-73. [doi: 10.1521/aeap.2006.18.supp.59] [Medline: 16987089]
38. Gale RC, Wu J, Erhardt T, Bounthavong M, Reardon CM, Damschroder LJ, et al. Comparison of rapid vs in-depth qualitative analytic methods from a process evaluation of academic detailing in the Veterans Health Administration. Implement Sci 2019 Feb 01;14(1):11 [FREE Full text] [doi: 10.1186/s13012-019-0853-y] [Medline: 30709368]
39. BugHerd. BugHerd.com.: Splitrock Studio URL: bugherd.com [accessed 2020-03-19]
40. Brown W, Yen P, Rojas M, Schnall R. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
41. Lau M, Lin H, Flores G. Racial/Ethnic Disparities in Health and Health Care among U.S. Adolescents. Health Serv Res 2012 Mar 14;47(5):2031-2059. [doi: 10.1111/j.1475-6773.2012.01394.x]
42. www.techsmith.com/tutorial-morae-current.html.: TechSmith Corporation URL: https://www.techsmith.com/morae-features.html [accessed 2020-03-19]
43. Yen PY, Wantland D, Bakken S. Development of a Customizable Health IT Usability Evaluation Scale. 2010 Presented at: AMIA Annu Symp Proc; 2010; Washington, DC p. 917-921.
44. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
45. Lewis JR. Psychometric Evaluation of the Post-Study System Usability Questionnaire: The PSSUQ. In: Proceedings of the Human Factors Society Annual Meeting. Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. Proceedings of the Human Factors Society 36th Annual Meeting; 1992 Presented at: Human Factors Society Annual Meeting; 1992; Boca Raton, FL p. 1259-1260. [doi: 10.1177/154193129203601617]
46. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979;2(3):197-207. [Medline: 10245370]
47. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
48. Saldaña J. The coding manual for qualitative researchers. 3rd ed. Los Angeles: Sage Publications; 2015.
49. Arnold EA, Sterrett-Hong E, Jonas A, Pollack LM. Social networks and social support among ball-attending African American men who have sex with men and transgender women are associated with HIV-related outcomes. Glob Public Health 2018 Feb;13(2):144-158 [FREE Full text] [doi: 10.1080/17441692.2016.1180702] [Medline: 27169632]
50. Phillips G, Peterson J, Binson D, Hidalgo J, Magnus M, YMSM of color SPNS Initiative Study Group. House/ball culture and adolescent African-American transgender persons and men who have sex with men: a synthesis of the literature. AIDS Care 2011 Apr;23(4):515-520. [doi: 10.1080/09540121.2010.516334] [Medline: 21271403]
51. nodejs.org/en/about/.: OpenJS Foundation URL: nodejs.org/en/about/ [accessed 2020-03-19]
52. mongodb.com/company.: MongoDB, Inc URL: mongodb.com/company [accessed 2020-03-16]
53. vuejs.org/v2/guide/.: Vue.js URL: vuejs.org/v2/guide/ [accessed 2020-03-18]
54. King D, Greaves F, Exeter C, Darzi A. 'Gamification': influencing health behaviours with games. J R Soc Med 2013 Mar;106(3):76-78. [doi: 10.1177/0141076813480996] [Medline: 23481424]
55. Bakkes S, Tan C, Pisan Y. Personalised gaming: a motivation and overview of literature. In: Proceedings of The 8th Australasian Conference on Interactive Entertainment: Playing The System. 2336731; 2012 Presented at: The 8th Australasian

Conference on Interactive Entertainment: Playing The System; 2012; Auckland, New Zealand p. 1-10. [doi: 10.1145/2336727.2336731]
56. Tabak M, Dekker-van WM, van DH, Vollenbroek-Hutten M. Promoting Daily Physical Activity by Means of Mobile Gaming: A Review of the State of the Art. Games Health J 2015 Dec;4(6):460-469. [doi: 10.1089/g4h.2015.0010] [Medline: 26397179]
57. Holloway IW, Jordan SP, Dunlap SL, Ritterbusch A, Reback CJ. Leveraging Social Networks and Technology for HIV Prevention and Treatment With Transgender Women. AIDS Educ Prev 2020 Apr;32(2):83-101 [FREE Full text] [doi: 10.1521/aeap.2020.32.2.83] [Medline: 32539480]
58. Sun CJ, Anderson KM, Kuhn T, Mayer L, Klein CH. A Sexual Health Promotion App for Transgender Women (Trans Women Connected): Development and Usability Study. JMIR Mhealth Uhealth 2020 May 12;8(5):e15888 [FREE Full text] [doi: 10.2196/15888] [Medline: 32396131]
59. Sun CJ, Anderson KM, Mayer L, Kuhn T, Klein CH. Findings from Formative Research to Develop a Strength-Based HIV Prevention and Sexual Health Promotion mHealth Intervention for Transgender Women. Transgend Health 2019;4(1):350-358 [FREE Full text] [doi: 10.1089/trgh.2019.0032] [Medline: 32042925]
60. Tanner AE, Mann-Jackson L, Song EY, Alonzo J, Schafer KR, Ware S, et al. Supporting Health Among Young Men Who Have Sex With Men and Transgender Women With HIV: Lessons Learned From Implementing the Intervention. Health Promot Pract 2020 Sep;21(5):755-763. [doi: 10.1177/1524839920936241] [Medline: 32757841]

Abbreviations

CSQ: Client Satisfaction Questionnaire
Health-ITUES: Health Information Technology Usability Evaluation Scale
mHealth: mobile health
NIH: National Institutes of Health
PEP: postexposure prophylaxis
PrEP: pre-exposure prophylaxis
PSSUQ: Post-Study System Usability Questionnaire
RCT: randomized controlled trial
TasP: Treatment as Prevention
YTW: young transgender women

Edited by R Kukafka; submitted 02.07.20; peer-reviewed by C Biernesser, T Kuhn; comments to author 04.08.20; revised version received 28.09.20; accepted 08.02.21; published 31.03.21.

Please cite as:
Kuhns LM, Hereth J, Garofalo R, Hidalgo M, Johnson AK, Schnall R, Reisner SL, Belzer M, Mimiaga MJ
A Uniquely Targeted, Mobile App-Based HIV Prevention Intervention for Young Transgender Women: Adaptation and Usability Study
J Med Internet Res 2021;23(3):e21839
URL: https://www.jmir.org/2021/3/e21839
doi:10.2196/21839
PMID:33787503

Original Paper

Comparative Success of Recruitment Strategies for an Exercise Intervention Trial Among Women With Polycystic Ovary Syndrome: Observational Study

Jamie L Benham[1,2], MD; Jane E Booth[2], MSc; Christine M Friedenreich[2,3,4,5], PhD; Doreen M Rabi[1,2,6], MD, MSc; Ronald J Sigal[1,2,5,6], MD, MPH

[1]Department of Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[2]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[3]Department of Cancer Epidemiology and Prevention Research, Cancer Care Alberta, Alberta Health Services, Calgary, AB, Canada

[4]Department of Oncology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[5]Faculty of Kinesiology, University of Calgary, Calgary, AB, Canada

[6]Department of Cardiac Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Ronald J Sigal, MD, MPH
Department of Medicine
Cumming School of Medicine
University of Calgary
Richmond Road Diagnostic and Treatment Centre, Room 1898
1820 Richmond Road SW
Calgary, AB, T2T 5C7
Canada
Phone: 1 403 955 8327
Email: rsigal@ucalgary.ca

Abstract

Background: Effective and efficient participant recruitment is a key determinant of the success of a research program. Previously reported recruitment strategies have displayed variable success rates in studies on women with polycystic ovary syndrome (PCOS).

Objective: This study aimed to evaluate the effectiveness and cost per participant of the recruitment strategies that we used in a prospective randomized controlled trial to examine the effects of exercise training among inactive women with PCOS, who are aged 18-40 years.

Methods: The 4 recruitment methods we used were as follows: (1) referral by health care providers or by word of mouth, (2) media (eg, local newspaper stories and radio interviews), (3) Facebook advertisements, and (4) unpaid advertisements including posters and websites. The proportions of potential, eligible, and enrolled participants recruited with each method were determined and compared using tests of proportion. The time investment and cost per participant enrolled were calculated for each recruitment strategy.

Results: Of 200 potential participants screened, 98 (49%) were recruited from unpaid advertisements (posters and websites), 70 (35%) from Facebook advertisements, 16 (8%) by referral, and 16 (8%) from traditional media (newspaper and radio). Every potential participant was recruited from separate means (ie, no participant was approached through more than one recruitment method). A total of 109 (54.5%) women were deemed eligible for participation in the trial, and 60 (30.0%) were enrolled. The proportion of potential participants who completed the trial was higher for those recruited from traditional media than from Facebook advertisements (n=7/16, 44% vs n=13/70, 19%, respectively; *P*=.03) or unpaid advertisements (n=7/16, 44% vs n=13/98, 13%, respectively; *P*=.002). The cost per participant was Can $18.21 (US $14.46) for Facebook advertisements and Can $43.88 (US $34.85) for unpaid advertisements. There were no direct trial costs for referrals or traditional media.

Conclusions: For this trial, each method was important for recruiting inactive women with PCOS because no participant reported learning about the trial through more than one method. Unpaid advertisements and Facebook advertisements helped recruit the largest number of participants in the trial, the former resulting in a higher cost per participant than the latter.

Trial Registration: ClinicalTrials.gov NCT03362918; https://clinicaltrials.gov/ct2/show/NCT03362918

(J Med Internet Res 2021;23(3):e25208) doi:10.2196/25208

KEYWORDS

aerobic exercise; exercise; exercise training; ovary; polycystic ovary syndrome; recruitment; well-being; women's health

Introduction

Polycystic ovary syndrome (PCOS) is a common endocrine disorder that affects up to 1 in 5 women of reproductive age [1]. Lifestyle interventions including exercise training and dietary modifications are encouraged for the management of PCOS [2], yet the optimal exercise prescription is unknown [3]. To address this issue, we designed a randomized controlled trial to evaluate the effect of 2 forms of exercise training—high-intensity interval training (HIIT) and continuous aerobic exercise training (CAET)—on the reproductive, anthropometric, and cardiometabolic outcomes of women of reproductive age living with PCOS compared to those in a no-exercise control group.

The ability to recruit participants effectively while simultaneously meeting the inclusion and exclusion criteria of a research program in a timely manner is a key determinant of the trial's ultimate success. Previous trials among women with PCOS employed various strategies to recruit participants, including referrals by health care providers, web-based advertisements, posters, and Facebook advertisements, all of which have displayed variable effectiveness [4-6]. Qualitative studies have reported that advertising on social media platforms and posting flyers at gynecology clinics might be effective strategies to recruit women living with PCOS to participate in a research study [7]. While advertising on social media platforms has been identified as a potentially effective recruitment strategy, the relative cost per participant from using Facebook advertisements to recruit women with PCOS has not been compared with that of more traditional strategies such as physician referrals and paid advertisements.

This study aimed to compare the success of 4 strategies used to recruit physically inactive women of reproductive age living with PCOS to a behavioral intervention trial.

Methods

Study Design

We conducted a single-center prospective randomized controlled trial to assess the effect of 2 forms of exercise training—high-intensity interval training (HIIT) and continuous aerobic exercise training (CAET)—on the reproductive, anthropometric, and cardiometabolic health markers in women with PCOS compared to those in a no-exercise control group [8]. The study period was 15 months and comprised 3 phases: (1) a 3-month run-in phase to assess baseline reproductive function; (2) a 6-month intervention phase where participants were randomized into HIIT, CAET, or no-exercise control groups; and (3) a 6-month follow-up phase. The run-in and intervention phases required daily assessment of menstrual status and ovulation. During the intervention phase, the participants randomly assigned to exercise training were prescribed 3 exercise sessions per week. The Conjoint Health Research Ethics Board of the University of Calgary and Alberta Health Services (REB17-1574) approved the study, and all participants provided written informed consent.

Participants

Physically inactive women with PCOS between the ages of 18 and 40 years were recruited in this study. The exclusion criteria were as follows: (1) taking medication that interferes with ovulation (eg, estrogens, progestins, glucocorticoids, metformin, gonadotropins, clomiphene citrate, or letrozole), (2) participating in regular exercise training for >40 minutes per week, and (3) having medical conditions that prevent them from receiving exercise training. Potential participants were initially screened through a telephone call on the basis of the inclusion and exclusion criteria and then through screening investigations including blood tests and electrocardiography. On meeting all trial inclusion and exclusion criteria, potential participants were invited for an in-person assessment and enrolled in the trial.

Recruitment Strategies

Recruitment for the trial was carried out between December 2017 and December 2018 in Calgary (Alberta, Canada) through 4 recruitment strategies simultaneously: Facebook advertisements, unpaid advertisements, traditional media, and referrals.

Facebook Advertisements

In total, 3 paid Facebook advertisements were used during trial recruitment. Each advertisement featured a cropped image of an individual performing exercise training and was captioned with the following phrase: "Can exercise help with Polycystic Ovary Syndrome (PCOS)?" with a click-through to a landing page that featured details on trial eligibility. The advertisements were consistent with Facebook's advertising policy [9]. The advertisements were targeted in accordance with location (ie, Calgary) and demographics (ie, women aged 18-40 years). We did not target advertisements on the basis of interests, behaviors, or connections. The first 2 advertisements were run for 2 weeks each in April and July 2018 (Figure 1) and the final advertisement was run for 1 week in September 2018.

Figure 1. Facebook advertisement run in July 2018.

Unpaid Advertisements

We used printed posters and websites. The printed posters featured an image of several individuals cycling indoors, and the title read: "Do you have Polycystic Ovary Syndrome (PCOS)? Would you like to become more physically active?" They were displayed locally on public poster boards at various sites including coffee shops, grocery stores, university campuses, public libraries, and recreation centers. Information regarding the trial was also posted on the following websites: (1) the University of Calgary research services office website [10]; (2) the O'Brien Institute of Public Health, Cumming School of Medicine, University of Calgary website [11]; and (3) ClinicalTrials.gov [12].

Traditional Media

Several research team members involved in the trial were interviewed about the trial for a University of Calgary publication "UToday." A member of the research team participated in broadcast interviews about the trial on 2 local radio stations.

Referrals

Local endocrinologists, gynecologists, family physicians, personal trainers, dieticians, and naturopaths were provided with information about the trial to disseminate to their patients who might meet the eligibility criteria of the trial. Participants were also recruited through word of mouth.

Recruitment Costs

Over the course of recruitment for the trial, costs incurred directly by the trial for each recruitment strategy were recorded. The time that trial team members spent on each of the recruitment methods was approximated and multiplied by the hourly compensation rate of the research assistant (Can $25 [US $19.86]). The cost of poster printing was estimated.

Statistical Analysis

The number of potential participants recruited through each recruitment method was determined. The proportions of potential participants who were eligible for enrollment, enrolled, randomized, and completed the trial was calculated for each recruitment method (Table 1). Tests of proportion were used to compare between-group differences. The time investment per enrolled participant was calculated by dividing the total number of hours spent on a recruitment strategy by the number of participants enrolled using that strategy. For each recruitment strategy, the cost per participant was determined by dividing the total cost of the recruitment strategy by the total number of participants enrolled using that strategy. Statistical significance was set at $P<.05$. The analysis was conducted using STATA (version 15.1, StataCorp).

Table 1. Outcome measure definitions.

Outcome measure	Definition
Participants recruited	The number of potential participants recruited to the trial.
Proportion of participants eligible for enrollment	The number of potential participants who were eligible to be enrolled in the trial divided by the total number of participants recruited.
Enrolled in the trial	The number of participants who were enrolled in the trial divided by the total number of participants recruited.
Randomized	The number of participants who completed the 3-month run-in phase and then randomized divided by the total number of participants recruited.
Completed the trial	The number of participants who completed the 9-month trial divided by the total number of participants recruited.
Time investment per participant	The total number of hours spent on a recruitment strategy divided by the number of participants enrolled using that strategy.
Cost per participant	The total cost of a recruitment strategy divided by the total number of participants enrolled using that strategy.

Results

Results Summary

Of the 200 potential participants screened for this pilot trial, 98 (49%) were recruited from unpaid advertisements, 70 (35%) from Facebook advertisements, 16 (8%) from traditional media, and 16 (8%) through referrals (Table 2). We observed no between-group differences in the proportion of participants who were eligible for enrollment or those who were enrolled in the trial. The proportion of potential participants who completed the 3-month run-in phase and then randomized was significantly greater for those recruited from traditional media than for those recruited from Facebook advertisements (n=7/16, 44% vs n=5/70, 24%, respectively; *P*=.03). The proportion of potential participants who completed the trial was significantly greater for those recruited from traditional media than for those recruited from Facebook advertisements (n=7/16, 44% vs n=13/70, 19%, respectively; *P*=.03) or unpaid advertisements (n=7/16, 44% vs n=13/98, 13%, respectively; *P*=.002). No participants reported learning about the trial from more than 1 recruitment method.

Table 2. Recruitment and completion of an exercise trial for women with polycystic ovary syndrome in Calgary (Alberta, Canada) in 2017-2020 (N=60).

Condition	All recruitment methods	Unpaid advertisements[a]	Facebook advertisements	Traditional media[b]	Referral[c]
Recruited, n	200	98	70	16	16
Eligible for enrollment, n (%)	109 (55.0)	49 (50)	39 (56)	8 (50)	10 (63)
Enrolled in the trial, n (%)	60 (30.0)	25 (26)	21 (30)	7 (44)	7 (44)
Randomized, n (%)	47 (24.0)	19 (19)	15 (21)	7 (44)	6 (38)
Completed the trial, n (%)	37 (19.0)	13 (13)	13 (19)	7 (44)	4 (25)
Age of the enrolled participants (years), mean (SD)	29.5 (4.9)	29.2 (5.2)	31.0 (4.4)	27.2 (5.5)	27.8 (3.1)
Employed, n (%)	54 (90.0)	23 (92)	18 (86)	6 (86)	7 (100)
Had a partner, n (%)	37 (62.0)	18 (72)	15 (71)	2 (29)	2 (29)
Had ≥1 child, n (%)	9 (15.0)	3 (12)	5 (24)	1 (14)	0 (0)
Current smoker, n (%)	6 (10.0)	4 (16)	2 (10)	0 (0)	0 (0)
Former smoker, n (%)	16 (27.0)	7 (28)	7 (33)	1 (14)	1 (14)

[a]Unpaid advertisements included printed posters and websites.

[b]Traditional media included radio interviews and a published interview.

[c]Referrals included referrals from health care providers and word of mouth.

Cost of Recruitment

For Facebook advertisements, the research team members spent approximately 15 hours in preparing the advertisements and liaising with Facebook personnel for advertisement approval. The time investment was 0.2 hours per enrolled participant. In terms of cost, each of the 3 advertisements cost Can $300 (US $238.30), adding up to a total of Can $900 (US $714.89), and the time-based compensation per research assistant was Can $375 (US $297.87; 15 hours at Can $25 [US $19.86] per hour). The cost per participant enrolled in the trial from Facebook advertisements was Can $18.21 (US $14.46).

Costs for unpaid advertisements were incurred for the time spent by the research team on developing posters (~10 hours at Can $25 [US $19.86] per hour), distributing them (~150 hours at Can $25 [US $19.86] per hour), and printing them (Can $300 [US $238.30]). The time investment per enrolled participant was 1.6 hours. The cost per participant enrolled in the trial from unpaid advertisements was Can $43.88 (US $34.85).

The trial incurred no direct costs for traditional media and referrals. For recruiting participants from traditional media, a study investigator (JLB) spent 3 hours. The time investment per enrolled participant was 0.2 hours. For referrals, the time investment was approximately 0.1 hours per enrolled participant for 2 hours invested by the study investigators.

Discussion

Principal Findings

A combination of 4 strategies facilitated successful participant recruitment in this pilot randomized controlled trial among women of reproductive age living with PCOS. Each of these methods was useful in meeting the participant recruitment target. Unpaid advertisements and Facebook advertisements yielded the greatest number of potential participants, while traditional media was most successful in recruiting participants who completed the trial. The cost per enrolled participant and the time investment by members of the research team was lower for Facebook advertisements than for unpaid advertisements.

The published data on the effectiveness of recruitment strategies for women with PCOS are conflicting. One study examining the effect of diet on weight loss and the risk of endometrial cancer in women with PCOS only recruited 11 of 40 desired participants through multiple recruitment strategies including flyers posted in a specialty clinic, referrals by health care providers, and the inclusion of trial information on a clinic's website [4]. Furthermore, Pastore and Dalal [6] evaluated the effect of acupuncture in 70 women living with PCOS recruited from among 89 participants as their estimated cohort size, in 36 months. They reported that continual enrollment required multiple recruitment strategies including posters, emails, website listings, direct mail, physician referrals, and traditional advertising in the form of TV, radio, and newspapers. Similarly, we found that multiple concurrent recruitment strategies were necessary to achieve our recruitment target.

Facebook advertisements were used successfully as a recruitment strategy in a previous trial involving women with PCOS [5], which used several recruitment strategies including Facebook advertisements, online advertisements, referrals, and flyers. Of these, Facebook advertisements yielded the highest proportion of potential participants. In our experience, Facebook advertisements were effective in recruiting potential trial participants in 5 weeks of running the advertisements; however, they recruited a lower proportion of potential participants than unpaid advertisements. A recent systematic review [13] examining the role of Facebook advertisements in participant recruitment reported that when compared with traditional recruitment methods including the television, radio, and print media, Facebook advertisements had shorter recruitment periods and lower costs and may have allowed for the inclusion of hard-to-reach demographics including young women.

In our trial, sole dependance on referrals from health care providers would not have been effective in recruiting the number of participants required for this trial. Over 12 months, only 16 potential participants were referred by health care providers. Women living with PCOS often only present for PCOS diagnosis and management if they are concerned by one or more symptoms such as difficulty losing weight, hirsutism, or infertility [14]. They may also prefer immediate treatment for those symptoms and thus may be reluctant to engage in a clinical trial that excludes women taking medical therapy that may affect ovulation. These women may also be motivated to engage in a lifestyle program that involves exercise training as a part of active management of PCOS and might be less likely to enroll in a trial where they may randomly be assigned to a no-exercise control group for 6 months.

Strengths and Limitations

The notable strength of this study is its provision of detailed data on the comparative success of recruitment strategies used in a pilot randomized controlled trial involving women of reproductive age living with PCOS. There are several limitations to consider when reviewing our results. First, participants who were recruited through word of mouth were not asked specifically how the individual who informed them about the study learned about it. Thus, the reach of our other recruitment strategies (unpaid advertisements, Facebook advertisements, or traditional media) may have been underestimated. Second, we did not use paid traditional media advertisements (ie, television, radio, or newspapers) in our pilot trial, which may be used in a larger trial; therefore, we could not compare the relative success of our recruitment strategies with that of traditional paid media advertisements. Finally, we did not collect data on race, ethnicity, or socioeconomic status to explore the effectiveness of each recruitment strategy on the basis of these variables. Future studies are needed to explore the effect of participant recruitment strategies on participant diversity.

Conclusion

In conclusion, all 4 recruitment strategies used in this trial were important for recruiting women living with PCOS, which was evident from the finding that none of the participants was recruited through more than 1 strategy. Facebook advertisements and unpaid advertisements yielded the largest number of participants enrolled in the trial. Facebook advertisements were more cost-effective and required less time investment than unpaid advertisements. In future trials, multiple concurrent recruitment strategies should be used to increase participant recruitment.

Acknowledgments

We would like to express our gratitude to the participants of this clinical trial. We would like to acknowledge the work of the research staff of the Clinical Trials Unit, Cumming School of Medicine, University of Calgary, for their contribution in study coordination and data collection. This clinical trial was supported by a grant from the Cumming School of Medicine, University of Calgary/Alberta Health Services Clinical Research Fund as well as a grant from the Libin Cardiovascular Institute of Alberta. The funders had no role in study design; data collection, analysis, or interpretation; or manuscript production or decision to submit for publication. JLB was supported by a Dr. Fernand Labrie Fellowship Grant from the Canadian Society of Endocrinology and Metabolism.

Authors' Contributions

JLB, JEB, CMF, DMR, and RJS conceived and designed the study. JLB, JEB, and RJS analyzed the data. JLB drafted the manuscript. JLB, JEB, CMF, DMR, and RJS critically revised the manuscript and approved the final version for submission. JLB and RJS are the guarantors of the study and take responsibility for the integrity of the data.

Conflicts of Interest

None declared.

References

1. Ding T, Hardiman PJ, Petersen I, Wang F, Qu F, Baio G. The prevalence of polycystic ovary syndrome in reproductive-aged women of different ethnicity: a systematic review and meta-analysis. Oncotarget 2017 Nov 10;8(56):96351-96358 [FREE Full text] [doi: 10.18632/oncotarget.19180] [Medline: 29221211]
2. Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, International PCOS Network. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Hum Reprod 2018 Sep 01;33(9):1602-1618 [FREE Full text] [doi: 10.1093/humrep/dey256] [Medline: 30052961]
3. Legro RS, Arslanian SA, Ehrmann DA, Hoeger KM, Murad MH, Pasquali R, Endocrine Society. Diagnosis and treatment of polycystic ovary syndrome: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 2013 Dec;98(12):4565-4592 [FREE Full text] [doi: 10.1210/jc.2013-2350] [Medline: 24151290]
4. Atiomo W, Read A, Golding M, Silcocks P, Razali N, Sarkar S, et al. Local recruitment experience in a study comparing the effectiveness of a low glycaemic index diet with a low calorie healthy eating approach at achieving weight loss and reducing the risk of endometrial cancer in women with polycystic ovary syndrome (PCOS). Contemp Clin Trials 2009 Sep;30(5):451-456. [doi: 10.1016/j.cct.2009.05.001] [Medline: 19477300]
5. Vizza L, Smith CA, Swaraj S, Agho K, Cheema BS. The feasibility of progressive resistance training in women with polycystic ovary syndrome: a pilot randomized controlled trial. BMC Sports Sci Med Rehabil 2016;8:14 [FREE Full text] [doi: 10.1186/s13102-016-0039-8] [Medline: 27175282]
6. Pastore LM, Dalal P. Recruitment strategies for an acupuncture randomized clinical trial of reproductive age women. Complement Ther Med 2009 Aug;17(4):229-235 [FREE Full text] [doi: 10.1016/j.ctim.2009.03.004] [Medline: 19632551]
7. Ee C, Smith C, Costello M, MacMillan F, Moran L, Baylock B, et al. Feasibility and acceptability of a proposed trial of acupuncture as an adjunct to lifestyle interventions for weight loss in Polycystic Ovary Syndrome: a qualitative study. BMC Complement Altern Med 2018 Nov 08;18(1):298 [FREE Full text] [doi: 10.1186/s12906-018-2358-7] [Medline: 30409195]
8. Benham JL, Booth JE, Corenblum B, Doucette S, Friedenreich CM, Rabi DM, et al. Exercise Training and Reproductive Outcomes in Women with Polycystic Ovary Syndrome: A Pilot Randomized Controlled Trial. Clin Endocrinol (Oxf) 2021 Feb 27 [FREE Full text] [doi: 10.1111/cen.14452] [Medline: 33638879]
9. Advertising Policies. Facebook. URL: https://www.facebook.com/policies/ads/# [accessed 2020-09-20]
10. Participate in Research at UCalgary. University of Calgary Research Services Office. Canada: University of Calgary; 2019 Dec 31. URL: https://www.ucalgary.ca/research/participate/
11. Papers and Studies. University of Calgary. URL: https://obrieniph.ucalgary.ca/research/papers-and-studies [accessed 2020-10-13]
12. ClinicalTrials.gov. NIH: U.S. National Library of Medicine. URL: clinicaltrials.gov [accessed 2021-03-12]
13. Whitaker C, Stevelink S, Fear N. The Use of Facebook in Recruiting Participants for Health Research Purposes: A Systematic Review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
14. Kitzinger C, Willmott J. 'The thief of womanhood': women's experience of polycystic ovarian syndrome. Soc Sci Med 2002 Feb;54(3):349-361. [doi: 10.1016/s0277-9536(01)00034-x] [Medline: 11824912]

Abbreviations

CAET: continuous aerobic exercise training
HIIT: high-intensity interval training
PCOS: polycystic ovary syndrome

Edited by R Kukafka; submitted 22.10.20; peer-reviewed by S Butts, K Reuter; comments to author 28.11.20; revised version received 23.12.20; accepted 25.02.21; published 30.03.21.

Please cite as:
Benham JL, Booth JE, Friedenreich CM, Rabi DM, Sigal RJ
Comparative Success of Recruitment Strategies for an Exercise Intervention Trial Among Women With Polycystic Ovary Syndrome: Observational Study
J Med Internet Res 2021;23(3):e25208
URL: https://www.jmir.org/2021/3/e25208
doi:10.2196/25208
PMID:33783363

Original Paper

Using the Ensuring Quality Information for Patients Tool to Assess Patient Information on Appendicitis Websites: Systematic Search and Evaluation

Shahi Ghani[1], BSc, MSc, DHMSA; Ka Siu Fan[1], BSc; Ka Hay Fan[2]; Lorenzo Lenti[1], BSc; Dimitri Raptis[3], MSc, PhD, MD

[1]St George's, University of London, London, United Kingdom

[2]Imperial College London, London, United Kingdom

[3]Department of HPB Surgery and Liver Transplantation, Royal Free Hospital, London, United Kingdom

Corresponding Author:
Shahi Ghani, BSc, MSc, DHMSA
St George's, University of London
Cranmer Terrace
Tooting
London, SW17 0RE
United Kingdom
Phone: 44 7772 075720
Email: shahi92@hotmail.com

Abstract

Background: Appendicitis is a common surgical problem among the young adult population, who are likely to use the internet to obtain medical information. This information may determine the health-seeking behavior of an individual and may delay medical attention. Little is known regarding the quality of patient information on appendicitis on the internet, as this has not been previously studied.

Objective: The aim of our study was to identify the quality of information regarding appendicitis on websites intended for the public.

Methods: We conducted a systematic review of information on appendicitis available online using the following 4 search terms in google: “appendicitis,” “appendix,” “appendectomy,” and “appendicectomy”. The top 100 websites of each search term were assessed using the validated Ensuring Quality Information for Patients (EQIP) tool (score 0-36).

Results: A total of 119 websites met the eligibility criteria for evaluation. The overall median EQIP score for all websites was 20 (IQR 18-22). More than half the websites originated from the USA (65/119, 54.6%), and 45.4% (54/119) of all websites originated from hospitals, although 43% (23/54) of these did not mention qualitative risks from surgery. Incidence rates were only provided for complications and mortality in 12.6% (15/119) and 3.3% (4/119) of all websites, respectively.

Conclusions: The assessment of the quality and readability of websites concerning appendicitis by the EQIP tool indicates that most sites online were of poor credibility, with minimal information regarding complication rates and mortality. To improve education and awareness of appendicitis, there is an immediate need for more informative and patient-centered websites that are more compatible with international quality standards.

(J Med Internet Res 2021;23(3):e22618) doi:10.2196/22618

KEYWORDS

appendicitis; patient information; EQIP tool; quality; tool; surgery; online health information; internet; health-seeking; behavior; review

Introduction

In the modern era, the increasing accessibility and availability of information has promoted the internet as the primary source for patient information. The access to countless sources of information can cater to every need by providing jargon-free material for the wider public while making details available for those who seek in-depth knowledge. Thus, many patients search

online for medical information on their symptoms prior to consulting medical professionals, many of whom subsequently self-diagnose based on these online sources [1,2]. Consequently, access to online medical information is critical to their decision-making process. However, with many sites containing potentially irrelevant or incorrect information, their credibility and reliability may present a barrier against seeking early medical help [3]. Furthermore, the use of unreliable websites may undermine patient relationships with health care professionals; at best, trust in healthcare may be affected, and at worst, presentations delayed by misinformed self-diagnoses may lead to poorer outcomes [4].

With a lifetime incidence of 8%, appendicitis is an affliction that is common enough to be familiar to the general public [5]. Hence, appendicitis, a leading cause of acute abdominal pain, and appendicectomy, its treatment, are likely to be terms that are searched by patients seeking further information. It has been shown that through providing education via high quality information on appendicitis and its risks, awareness and outcomes may be improved [6,7]. In order to make informed and competent health decisions, safe, reliable, and easily accessible information is essential; thus, the variable quality of internet patient resources warrant an evaluation of the quality and readability of this medical information.

The original Ensuring Quality Information for Patients (EQIP) tool is a checklist of 20 items used to assess written health care information [8]. Various aspects are considered, such as clarity of information, quality of written work, and website design. The EQIP tool has been used to evaluate information sources related to gallstone disease, transplant surgeries, eczema, liposuction, and, more recently, COVID-19 [9-13], demonstrating its applications across various disciplines and information types. We assessed the top-indexed websites related to appendicitis and appendicectomy using the modified EQIP tool for evaluation. The objective of our study was to evaluate the quality of information found on the top-searched websites that aim to provide patients with information on appendicitis.

Methods

Eligibility Criteria, Information Sources, and Data Selection

The most popular search engine, Google [14-16], was used to obtain a database of websites. Other search engines were not used in this study, as this would only lead to duplicate results. The searched terms were “appendix,” “appendicitis,” “appendicectomy”, and “appendectomy”. These were obtained using the Google AdWords Keyword Planner [17]. Formation of the database, analysis, and eligibility assessment of the websites were performed between September 2019 and February 2020. Previous work has suggested that patients limit their searches to well within the first 100 hits; therefore, the hits on each page were obtained until this target was reached [10]. The inclusion criterion was any website with information intended for patients. Websites were excluded if the literature was intended for scholars in scientific journals or if they were in a language other than English. Furthermore, links which directed individuals to purely video content or which were used for marketing purposes were also excluded. Upon exclusion, 119 websites were identified as eligible for analysis.

Website Scraping

To obtain the URL database from the top 100 hits for each search engine, a website scraping tool was developed. This reduced the amount of time required for cutting and pasting links to the database. The custom Hypertext Preprocessor (The PHP Group) tool was designed to make HTTP requests to the search engines, mimicking the requests web browsers make when using a search engine. This tool made repeated requests, logging all of the hits per page with a target of 100 unique URLs. During this process, any duplicates were automatically removed within the individual search. If there were more hits on a page after the target of 100 websites was reached, these websites were also collected. The tool was run using a server based in Texas in the United States, although no preferences were chosen to limit searches to certain geographical areas.

Data Entry

Each website was assessed independently by 4 assessors, SAG, KSF, KHF, and LL, all of whom are fluent in English. To assess each website, a Google Form containing the 36 EQIP items was used to evaluate criteria through yes, no, or “N/A” responses. Assessors also recorded the country of origin and the following source types: academic center, encyclopedia, health department, hospital, industry, news service, patient group, practitioner, professional society, or other. After the initial round of data entry, the websites were reassessed by another assessor, and any contradictory results were resolved by consensus.

EQIP Tool

The original EQIP tool has been expanded to 36 criteria to provide a more robust and effective analysis of patient information. The modified EQIP tool sets out to satisfy the patient information collaboration guidelines of both the British Medical Association (BMA) [18] and the International Patient Decision Aids Standards (IPDAS) [19]. The modified EQIP tool consists of 36 items split into 3 domains: content (items 1-18), identification (19-24), and structure (items 25-36). Similar to previous studies, only yes or no options were provided for each item to avoid assessor subjectivity in partial answers. The option for “N/A” was also included if items were not relevant for the type of source. Websites which scored above the 75th percentile were deemed high-scoring websites.

Additional Items Describing Mortality and Complication Rates of Surgery and Emergency Information

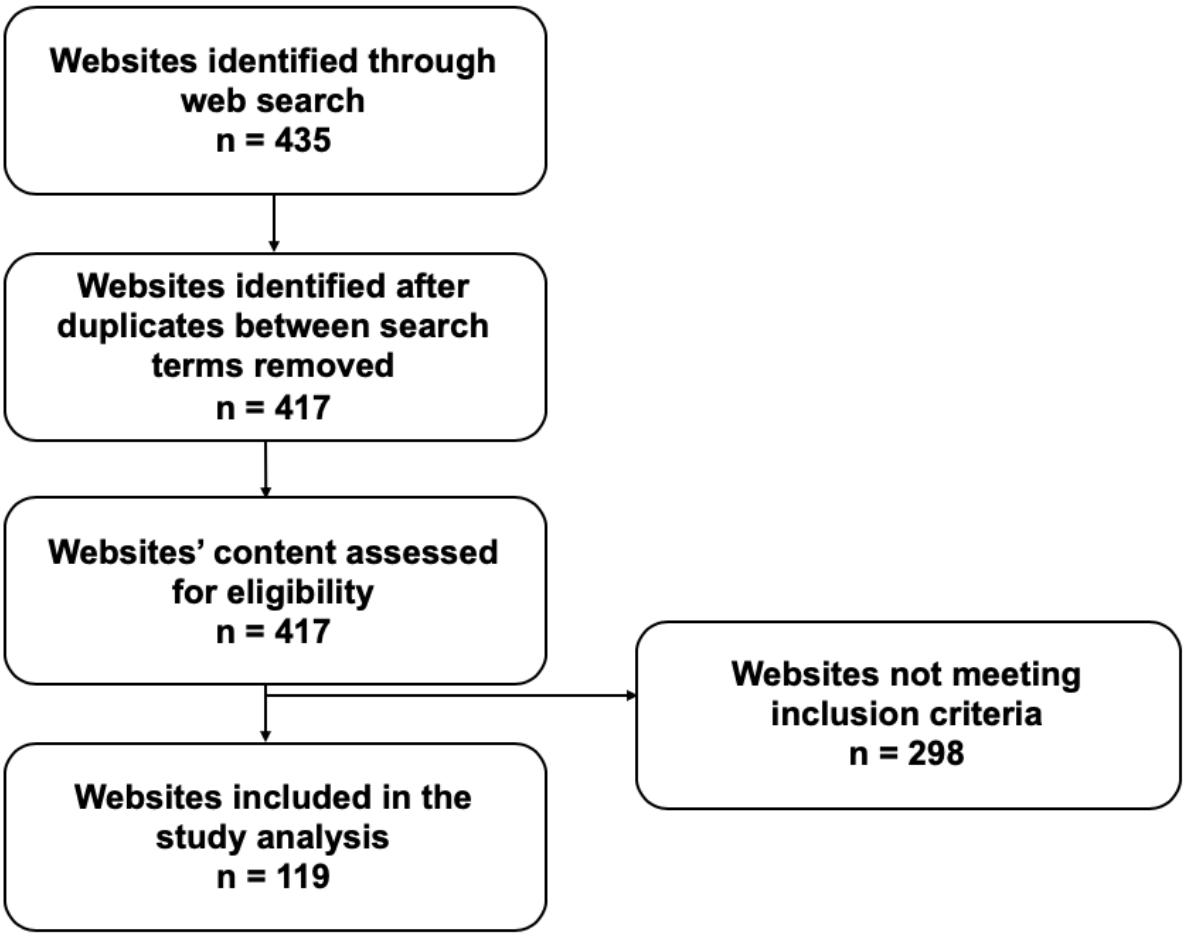
Questions were added to the Google Form to assess the variation in reported complication and mortality rates published by differing sources. The additional questions identified those websites that included rates for mortality and complications, and recorded the values given. Furthermore, a question was added to identify websites that included advice in the case of an emergency.

Statistical Analysis

Continuous variables are reported as median and IQR and categorical variables as numbers and proportions in percentages.

Continuous variables were compared with the Mann-Whitney and Kruskal-Wallis tests where appropriate. Proportions were compared with the Fisher exact test or chi-square test where appropriate. All *P* values were 2-sided and considered statistically significant when $P<.05$. A decision was made to dichotomize the EQIP score by using the 75th percentile as a cutoff point for discriminating high-scoring from low-scoring websites, as previously described and defined [9]. Statistical analysis was performed using R version 3.3.2 (The R Project for Statistical Computing, GNU General Public License version 2) and R Studio version 1.0.44 (RStudio) with the graphical user interface, rBiostatistics.com alpha version [20].

Figure 1. Workflow for identification of websites eligible for analysis.



Results

Gathering of Websites With Information on Appendicitis and Its Management

To obtain a database of websites for analysis, the unique hits from each page using the search terms ("appendix," "appendicitis," "appendicectomy," and "appendectomy") were gathered. The workflow of this is shown in Figure 1. Although a target of 100 websites per term was used, additional hits on the last page of each search were gathered if they were unique, resulting in a total of 435 websites. Duplicate results obtained between search terms and websites failing to meet inclusion criteria were removed, resulting in 119 websites for analysis.

Overall Quality of the Websites According to the Modified EQIP Tool

The distribution in EQIP score between all websites meeting inclusion criteria is shown in Figure 2. The country of origin and source of information for the database are shown in Table 1. Hospitals were the most common source of information, accounting for 45.4% (54/119) of the database, and 52% (28/54) of websites by hospitals originated from the United States. The country with the most websites was the United States, representing 54.6% (65/119) of the total, of which 26% (17/65) were high scoring. This represented 61% (16/26) of the total high-scoring websites. The distribution of EQIP scores by country of origin with more than one website is shown in Figure 3. Websites from the United Kingdom demonstrated the greatest variance in scores ranging from 9-28 with a median of 18 (Figure 3). Overall EQIP scores from countries with only 1 website were 24 for New Zealand, 19 for South Africa, 16 for Singapore, and 15 for India.

Figure 2. EQIP score of all websites assessed with patient information regarding appendicitis. High-scoring websites (EQIP score >75th percentile) are indicated by darker shadowing and labeling. EQIP: Ensuring Quality Information for Patients.

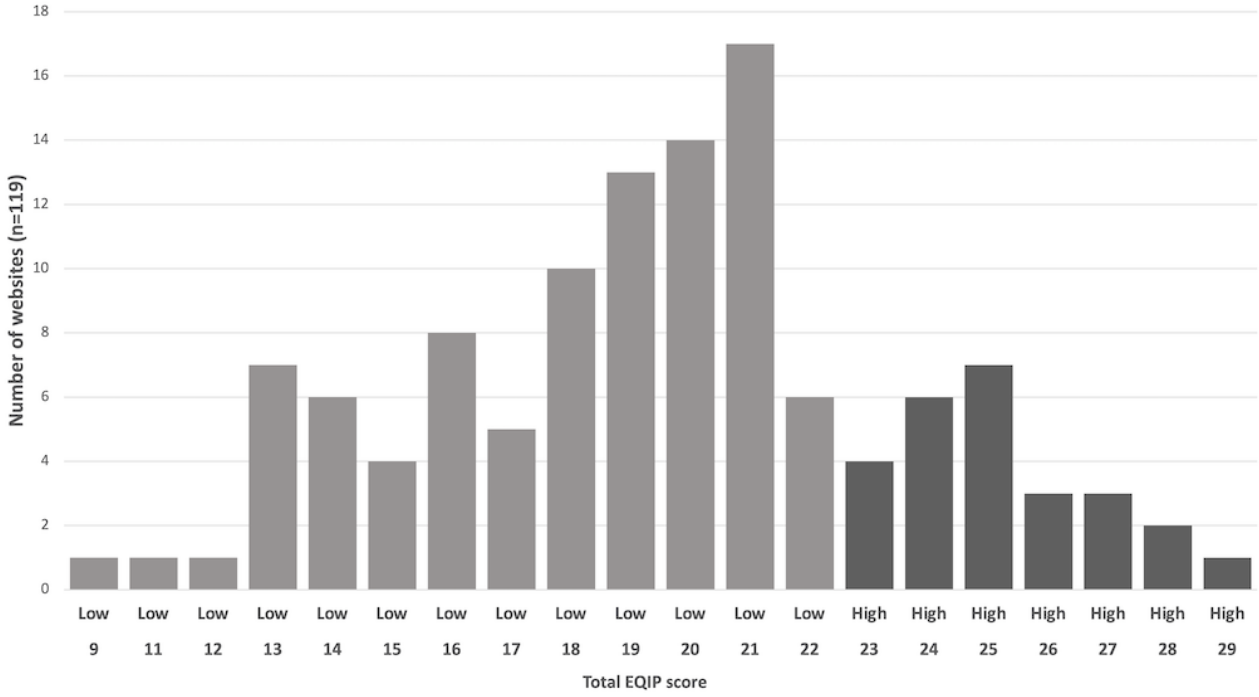
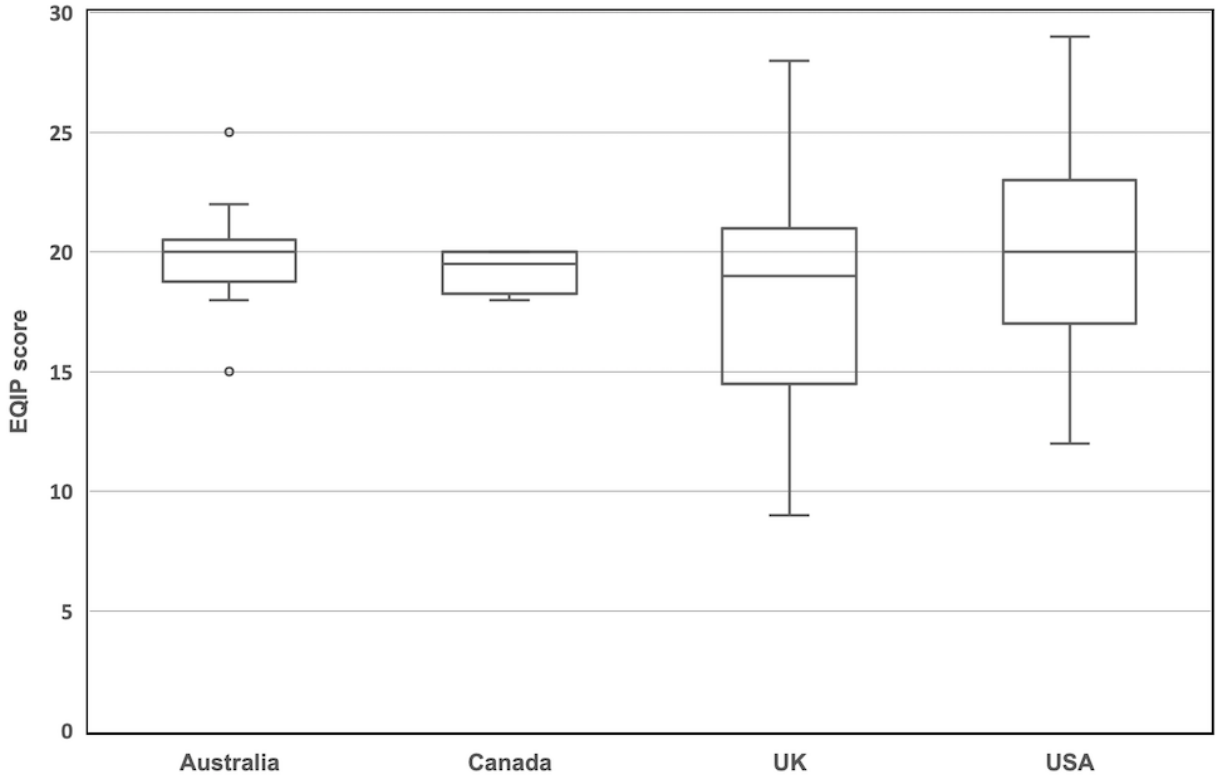


Table 1. Descriptive analysis of websites included in the study grouped by country of origin and source of information (N=119).

Parameters	Articles, n (%)
Country	
Australia	10 (8.4)
Canada	4 (3.4)
India	1 (0.8)
New Zealand	1 (0.8)
Singapore	1 (0.8)
South Africa	1 (0.8)
United Kingdom	36 (30.3)
United States	65 (54.6)
Source of information	
Academic center	5 (4.2)
Encyclopedia	6 (5.0)
Industry	14 (11.8)
Health department	12 (10.1)
Hospital	54 (45.4)
News service	19 (16.0)
Patient group	1 (0.8)
Professional society	8 (6.7)

Figure 3. Distribution of EQIP scores per country of origin with more than one website. Boxplots represent median (within box) and IQR (lower and upper lines). EQIP: Ensuring Quality Information for Patients.



Complication rates were included in 12.6% (15/119) of websites and varied between 0.2%-26%. Mortality rates were included in 3.3% (4/119) of websites and ranged between 0.001%-1.8%. Emergency advice was provided in 44.5% (53/119) of websites.

EQIP Content Data

The items for the content domain of the EQIP tool are shown in Table 2. The median score achieved was 9 (50%), and the maximum score obtained was 16 (89%) of a possible 18 (Table 3). All of the high-scoring websites provided a description of the medical problem (item 3), definition of the purpose of intervention (item 4), and description of the qualitative risks and complications of appendicectomy (item 6; Table 2). High-scoring websites were found to describe how complications are handled (item 12) and provided details of other sources of reliable information (item 17) in 77% (20/26) and 73% (19/26) of cases, respectively; this was significantly less in the low-scoring websites (P<.001; Table 2). IQRs for the content domain are included in Table 3.

Among all websites (including low scoring), 95.7% (114/119) failed to address the costs and insurance issues related to appendicectomy (item 15). Furthermore, 97.4% (116/119) and 86.5% (103/119) of websites failed to describe the quantitative benefits (item 8) and risks (item 10) of appendicectomy, respectively (Table 2). Of note, 23 of the 54 hospital-published sources (43%) did not include any description of the qualitative risks and complications of surgery (item 9).

Table 2. Breakdown of results for the content domain (items 1-18) of the modified Ensuring Quality Information for Patients tool (N=119).

Content domain item	Websites scoring points for domain items, n (%)			OR[a]	95% CI	*P* value
	Overall (N=119)	High scoring (n=26)	Low scoring (n=93)			
1. Initial definition of which subjects will be covered	85 (71)	22 (85)	63 (68)	2.60	0.78-11.30	.14
2. Coverage of the previously defined subjects (N/A[b] if the answer is "no" for item 1)	90 (76)	23 (88)	67 (72)	N/A	N/A	.26
3. Description of the medical problem/treatment/procedure	115 (97)	26 (100)	89 (96)	N/A	N/A	.58
4. Definition of the purpose of the interventions	110 (92)	26 (100)	84 (90)	N/A	N/A	.20
5. Description of treatment alternatives (conservative management)	45 (38)	14 (54)	31 (33)	2.32	0.88-6.22	.69
6. Description of the sequence of the interventions and surgical procedure	96 (81)	25 (96)	71 (76)	7.66	1.12-331.81	.02
7. Description of the qualitative benefits for the patient	90 (76)	25 (96)	65 (70)	10.63	1.58-456.75	.004
8. Description of the quantitative benefits to the patient	3 (3)	2 (8)	1 (1)	7.49	0.38-455.58	.12
9. Description of the qualitative risks and complications	76 (64)	26 (100)	50 (54)	N/A	N/A	<.001
10. Description of the quantitative risks and complications	16 (13)	10 (38)	6 (6)	8.82	2.51-34.13	<.001
11. Addressing quality-of-life issues	82 (69)	24 (92)	58 (62)	7.15	1.61-66.23	.003
12. Description of how complications are handled	36 (30)	20 (77)	16 (17)	15.52	5.05-55.21	<.001
13. Description of the precautions that the patient may take	37 (31)	13 (50)	24 (26)	2.85	1.06-7.75	.03
14. Mention of warning signs that the patient may detect	80 (67)	24 (92)	56 (60)	7.82	1.76-72.15	.002
15. Addressing medical intervention costs and insurance issues	5 (4)	1 (4)	4 (4)	0.89	0.02-9.55	.99
16. Specific contact details for hospital services (N/A if not hospitals)	29 (24)	9 (35)	20 (22)	11.27	1.37-530.56	.01
17. Specific details of other sources of reliable information/support	41 (35)	19 (73)	22 (24)	8.56	2.98 - 27.49	<.001
18. Coverage of all relevant issues for the topic (summary item for all content criteria)	20 (17)	16 (62)	4 (4)	33.65	166.47-8.77	<.001

[a]OR: odds ratio.

[b]N/A: not applicable.

Table 3. Analysis of EQIP scores obtained for each domain and overall.

Statistic	Content data[a]	Identification data[b]	Structure data[c]	Overall EQIP[d,e]
Median	9	3	8	20
Minimum	3	0	3	9
Maximum	16	6	10	29
Quartile 1	8	2	7	18
Quartile 3	11	4	9	22
IQR	3	2	2	4

[a]Total possible score for content=18.

[b]Total possible score for identification=6.

[c]Total possible score for structure=12.

[d]Total possible score=36.

[e]EQIP: Ensuring Quality Information for Patients.

EQIP Identification

In the identification domain, the median score obtained was 3 (50%), and the maximum score obtained was 6 (100%; Table 3). Two websites, both of which were high scoring (overall score >75th percentile) obtained maximum points for this section. High-scoring websites were significantly better (58%, 15/26) than low-scoring websites (20%, 19/93) in providing a short bibliography of the evidence base for the information ($P<.001$; item 23; Table 4). IQRs for the identification domain are included in Table 3.

Furthermore, 97.4% (116/119) of all websites failed to include a statement about how patients were involved or consulted in the document's production (item 24), and 68.0% (81/119) did not explicitly provide names of the persons or entities that financed the document (item 22; Table 4).

Table 4. Breakdown of results of the identification domain (items 19-24) and structure domain (items 25-36) of the modified Ensuring Quality Information For Patients tool (N=119).

Item	Websites scoring points for domain items, n (%)			OR[a]	95% CI	*P* value
	Overall (N=119)	High scoring (n=26)	Low scoring (n=93)			
Identification domain						
19. Date of issue or revision	75 (63)	23 (88)	52 (56)	5.97	1.63-33.20	.002
20. Logo of the issuing body	117 (98)	26 (100)	91 (98)	N/A[b]	N/A	.99
21. Names of the persons or entities that produced the document	84 (71)	23 (88)	61 (66)	3.98	1.08-22.28	.03
22. Names of the persons or entities that financed the document	38 (32)	10 (38)	28 (30)	1.54	0.55-4.21	.35
23. Short bibliography of the evidence-based data used in the document	34 (29)	15 (58)	19 (20)	5.22	1.90-14.91	<.001
24. Statement about whether and how patients were involved/consulted in the document's production	3 (3)	2 (8)	1 (1)	7.49	0.38-55.58	.12
Structure domain						
25. Use of everyday language and explanation of complex words or jargon	107 (90)	24 (92)	83 (89)	1.44	0.28-14.42	.99
26. Use of generic names for all medications or products (N/A if no medications described)	55 (46)	20 (77)	35 (38)	1.14	0.06-70.72	.99
27. Use of short sentences (<15 words on average)	107 (90)	25 (96)	82 (88)	3.33	0.44-149.95	.46
28. Personal address to the reader	90 (76)	23 (88)	67 (72)	2.95	0.79-16.65	.12
29. Respectful tone	117 (98)	26 (100)	91 (98)	N/A	N/A	.99
30. Clear information (no ambiguities or contradictions)	117 (98)	26 (100)	91 (98)	N/A	N/A	.99
31. Balanced information on risks and benefits	64 (54)	24 (92)	40 (43)	15.59	3.53-143.74	<.001
32. Presentation of information in a logical order	116 (98)	26 (100)	90 (97)	N/A	N/A	.99
33. Satisfactory design and layout (excluding figures or graphs; see next item)	104 (87)	24 (92)	80 (86)	1.94	0.40-18.91	.52
34. Clear and relevant figures or graphs (N/A if absent)	34 (29)	10 (38)	24 (26)	1.85	0.30-20.62	.70
35. Inclusion of a named space for the reader's notes or questions	12 (10)	2 (8)	10 (11)	0.69	0.07-3.59	.99
36. Inclusion of a printed consent form contrary to recommendations (N/A if not from hospitals)	1 (1)	1 (4)	0 (0)	N/A	N/A	.19

[a]OR: odds ratio.

[b]N/A: not applicable.

EQIP Structure

The median score obtained for the structure domain was 8 (66%), and the maximum score obtained was 10 (83%) of a possible 12 (Table 3). All high-scoring websites used a respectful tone (item 29), presented clear information (item 30), and delivered information in a logical order (item 32). High-scoring websites were found to include balanced information on risks and benefits (item 31) more frequently (24/26, 92%) than low-scoring websites (40/93, 43%; Table 4). In addition, 89.9% (107/119) of all websites failed to include a named space for readers' questions, and 71.4% (85/119) failed

to include clear and relevant figures or graphs (item 34; Table 4). IQRs for the structure domain are included in Table 3.

Top 3 Websites According to the EQIP Tool

The websites scoring above the 99th percentile (EQIP score of 28) are shown in Table 5. The top-ranked resource, which scored 29 out of 36, was produced by the American College of Surgeons, displaying a comprehensive guide for patients to understand appendicitis and its management. Bupa health insurance and Medical News Today, each with a with an EQIP score of 28, were tied for the second ranking. All 3 sources provided information on symptoms, details of the procedure, preoperative and postoperative instructions, and guidance on complications.

Table 5. Websites scoring above the 99th percentile (EQIP score of 28).

Organization	Reference	Content data[a]	Identification data[b]	Structure data[c]	Overall EQIP[d,e]
American College of Surgeons	[21]	15	4	10	29
Medical News Today	[22]	16	4	8	28
Bupa	[23]	12	7	9	28

[a]Total possible score for content=18.

[b]Total possible score for identification=6.

[c]Total possible score for structure=12.

[d]Total possible score=36.

[e]EQIP: Ensuring Quality Information for Patients.

Discussion

Principal Findings

Based on our analysis, the quality of patient information regarding appendicectomy was shown to be of a moderate level, as reflected by the median overall score of 20 (IQR 18-22). Generally, websites tended to score well in the structure domain, reflected by the median score of 8 (IQR 7-9). This domain focuses on the ability of websites to display their information in a clear and logical manner. As technology is improving, it is becoming easier to produce websites and leaflets to a higher visual standard with minimal computer literacy. This presents a new challenge to patients with limited clinical understanding, as websites of poorer quality may appear similar to high quality sources of information. The increase in total EQIP score due to higher marks being obtained from the structure domain from the improvement of website quality is a phenomenon which has also been seen in regards to COVID-19 [13].

In the identification domain, high-scoring websites were better than low-scoring websites at providing a bibliography of the evidence base, potentially suggesting that these were written by individuals with experience in academia. It is therefore unsurprising that high-scoring websites more often included the names of the persons that produced the document. Contrastingly, the majority of websites (68.0%, 81/119) both high- and low-scoring, failed to include the names of the persons or entity that financed the document.

As appendicitis commonly affects younger individuals, it is possible that the decision to go to hospital is being made by a parent or guardian of the patient. This decision may be the result of a combination of factors, such as the severity of symptoms experienced by the patient, the health beliefs of the parent or guardian, and the quality of any information they may seek. Delays in presentation to hospital may increase the risk of complications, such as rupture; therefore, it is important that information regarding complications and warning signs is clearly described. Despite this clear need for quality, the scores achieved for the content domain were generally low. Only 67.2% (80/119) of websites mentioned warning signs to detect appendicitis, although 44.5% (53/119) of the websites did provide some form of advice in case of emergencies.

Hospitals contributed 45.4% (54/119) to the total database of websites, but 43% (23/54) of these failed to the specify qualitative risks of complications after surgery. It is also concerning that only 30.2% (36/119) of all websites provided a description of how complications may be handled. Over half of the websites (52%,28/54) from hospitals originated from the United States, possibly reflecting a degree of hesitancy by private hospitals to include information that may adversely affect customer decisions. This is further supported by the low number of websites reporting the incidence rate of complications. Although mortality rates are low for appendicitis, it is disappointing that this information was only provided by 4 websites, as this should be standard practice with the aim to fully inform patients.

To the best of our knowledge this is the only study to date to evaluate patient information regarding appendix surgery; therefore, it is not possible to evaluate how the EQIP tool compares to other scoring systems used for this condition. However, the EQIP tool has been described previously when evaluating the quality of information for gallstone disease (median EQIP score 15, IQR 13-18) [9], for clefts of the lip and palate (median EQIP score 19, IQR 16-22) [24], for bariatric surgery (median EQIP score 17, IQR 15-19) [25], for phalloplasty (median EQIP score 17.5, IQR 13-21) [26], for Dupuytren disease (median EQIP score 16, IQR 13-19) [27], for breast augmentation (median EQIP score 15, IQR 13-17) [28], donor information for living liver transplantation (median EQIP score 16, IQR 13-20) [10], and for COVID-19 (median EQIP score 18, IQR 15-20) [13]. The median score and IQR of this study are slightly higher than those in these studies. This may suggest that the general quality of information for appendicitis on the internet is higher than that of the previously studied diseases; however, this relatively high quality is likely

due to the points gained in the structure domain in comparison to previous studies. The maximum score obtained in this study was 29, which was only achieved by 1 website and was still considerably lower than the maximum possible score of 36.

The first and second highest EQIP scores were achieved by websites from the United States and the United Kingdom, respectively. The United Kingdom also had the lowest-scoring website and lowest median among countries with more than one website. This indicates that the high-scoring websites in the United Kingdom are diluted by a majority of poor quality websites. The IQR range of scores for websites originating from Canada and Australia were much smaller, suggesting that, although there are fewer websites, they are generally of a higher quality.

Limitations

There were a number of limitations for this study. Identification of search terms with Google AdWords Keyword Planner only provides commonly used search phrases by the wider public and may not be able to truly predict the search patterns of individuals seeking health information. Another limitation is that only websites in English language were evaluated; thus, the conclusions drawn might not be representative of patient websites in other languages. Furthermore, we have described the use of the EQIP in relation to appendicitis although the tool was not originally created for this specific purpose, and therefore this may be considered as a limitation. However, as the EQIP tool has previously been shown to be robust and effective for a number of surgical conditions [9,10,26,28], it is reasonable to expect this to also hold true for appendicitis. One may argue that there is no clear reason for why tools such as EQIP exist to evaluate patient information. The purpose of this tool is to enable a method of categorizing information so that we may learn what areas can be improved upon to produce patient information websites of higher quality. Our next step is to use the information gathered from this paper to design a website that will include information which has been found to be regularly omitted from previous work. Using solely the Google search engine introduced another limitation, as it is possible that the search engine had listed results not simply by popularity, but also by the geographical location of the requesting computer. Therefore, although geolocation features were disabled, the websites extracted could have still been centered around a particular location or continent, preventing a truly representative analysis of the top websites used globally. Finally, it is important to note that the findings of this study act as a snapshot of a particular point in time when the search was used; however, while search engine results do change over time, we consider the findings of this study to be representative of the information available to patients.

Conclusions

In conclusion, the internet has become an essential source of information for our society. Our study showed that despite a growing body of web-based resources on appendicitis and appendicectomy, the currently available websites are generally of poor quality and inform patients inadequately. Although the clinicians responsible for each patient can provide patients with important, accurate, and relevant clinical information, the ability to direct patients to trustworthy internet resources may lead to better patient education and long-term outcomes. Online information is playing an increasingly significant role in patients' attitude and can potentially affect willingness to accept and comply with medical advice [29]. With emergencies such as appendicitis, it is paramount to make emergency guidance widely available, especially when websites may be a patient's first point of contact for information. Health care professionals should strive to educate patients on how to navigate and appraise internet-based resources in order to access the highest quality of information. Many studies have also identified similar problems and have not acted upon them, thus highlighting the urgent need to establish high quality websites and specific information evaluation tools to ensure the optimal patient education for a disease as common as appendicitis.

Authors' Contributions

SAG and DAR contributed to the study conception and design; SAG and KSF conducted the acquisition of data; SAG, KSF, KHF, and LL analyzed and interpreted the data; SAG, KSF, and KHF drafted the manuscript; and SAG, KSF, KHF, LL, and DAR critically revised the manuscript.

Conflicts of Interest

None declared.

References

1. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
2. Ryan A, Wilson S. Internet healthcare: do self-diagnosis sites do more harm than good? Expert Opin Drug Saf 2008 May;7(3):227-229. [doi: 10.1517/14740338.7.3.227] [Medline: 18462181]
3. Semigran H, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
4. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015 Jan 13;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]

5. Jones M, Lopez R, Deppen J. Appendicitis. StatPearls 2019;23:-1. [Medline: 29630245]
6. Russell K, Rollins MD, Barnhart DC, Mone MC, Meyers RL, Skarda DE, et al. Facilitating behavior change and well-being to improve health outcomes:charge awareness affects treatment choice: prospective randomized trial in pediatric appendectomy. Ann Surg 2015 Jul;262(1):189-193. [doi: 10.1097/SLA.0000000000000885] [Medline: 25185471]
7. Gold DT, McClung B. Approaches to patient education: emphasizing the long-term value of compliance and persistence. Am J Med 2006 Apr;119(4 Suppl 1):S32-S37. [doi: 10.1016/j.amjmed.2005.12.021] [Medline: 16563940]
8. Moult B, Franck LS, Brady H. Ensuring quality information for patients: development and preliminary validation of a new instrument to improve the quality of written health care information. Health Expect 2004 Jun;7(2):165-175 [FREE Full text] [doi: 10.1111/j.1369-7625.2004.00273.x] [Medline: 15117391]
9. Raptis DA, Sinanyan M, Ghani S, Soggiu F, Gilliland JJ, Imber C. Quality assessment of patient information on the management of gallstone disease in the internet - a systematic analysis using the modified ensuring quality information for patients tool. HPB (Oxford) 2019 Dec;21(12):1632-1640 [FREE Full text] [doi: 10.1016/j.hpb.2019.03.355] [Medline: 31174998]
10. Melloul E, Raptis DA, Oberkofler CE, Dutkowski P, Lesurtel M, Clavien P. Donor information for living donor liver transplantation: where can comprehensive information be found? Liver Transpl 2012 Aug;18(8):892-900 [FREE Full text] [doi: 10.1002/lt.23442] [Medline: 22467198]
11. McCool ME, Wahl J, Schlecht I, Apfelbacher C. Evaluating written patient information for eczema in German: comparing the reliability of two instruments, DISCERN and EQIP. PLoS One 2015;10(10):e0139895 [FREE Full text] [doi: 10.1371/journal.pone.0139895] [Medline: 26440612]
12. Zuk G, Palma AF, Eylert G, Raptis DA, Guggenheim M, Shafighi M. Systematic review of quality of patient information on liposuction in the internet. Plast Reconstr Surg Glob Open 2016 Jun;4(6):e759 [FREE Full text] [doi: 10.1097/GOX.0000000000000798] [Medline: 27482498]
13. Fan KS, Ghani SA, Machairas N, Lenti L, Fan KH, Richardson D, et al. COVID-19 prevention and treatment information on the internet: a systematic analysis and quality assessment. BMJ Open 2020 Sep 10;10(9):e040487 [FREE Full text] [doi: 10.1136/bmjopen-2020-040487] [Medline: 32912996]
14. Top 15 most popular search engines. eBiz MBA. URL: http://www.ebizmba.com/articles/search-engines [accessed 2020-11-01]
15. Search EW. Google hits the billion monthly unique visitors mark. Search Engine Watch. URL: https://www.searchenginewatch.com/2011/06/23/google-hits-the-billion-monthly-unique-visitors-mark/ [accessed 2020-11-01]
16. Google. Google. URL: https://www.google.com/ [accessed 2020-11-01]
17. Google AdWords. Google. URL: https://ads.google.com/intl/en_uk/home/tools/keyword-planner/ [accessed 2020-11-01]
18. British MA. BMA patient information awards. British Medical Association. 2018. URL: https://www.bma.org.uk/library/patient-information-awards [accessed 2020-11-01]
19. IPDAS collaboration background document. International Patient Decision Aid Standards Collaboration. 2005. URL: http://ipdas.ohri.ca/IPDAS_Background.pdf [accessed 2020-11-01]
20. Cloud graphical user interface for R statistics and eLearning platform. rBiostatistics. URL: https://www.rbiostatistics.com/ [accessed 2020-11-01]
21. American COS. Appendicectomy. American College of Surgeons. URL: https://www.facs.org/~/media/files/education/patient%20ed/app.ashx [accessed 2020-11-01]
22. Rachell N. What to know about appendicectomy. Medical News Today. URL: https://www.medicalnewstoday.com/articles/323805 [accessed 2020-11-01]
23. Appendicitis. Bupa. URL: https://www.bupa.co.uk/health-information/childrens-health/appendicitis [accessed 2020-11-01]
24. Karamitros G, Kitsos N. Clefts of the lip and palate: is the internet a trustworthy source of information for patients? Int J Oral Maxillofac Surg 2018 Sep;47(9):1114-1120. [doi: 10.1016/j.ijom.2018.03.003] [Medline: 29622479]
25. Vetter D, Ruhwinkel H, Raptis DA, Bueter M. Quality assessment of information on bariatric surgery websites. Obes Surg 2018 May;28(5):1240-1247. [doi: 10.1007/s11695-017-2983-0] [Medline: 29110245]
26. Karamitros GA, Kitsos NA, Sapountzis S. Systematic review of quality of patient information on phalloplasty in the internet. Aesth Plast Surg 2017 Jul 11;41(6):1426-1434. [doi: 10.1007/s00266-017-0937-5] [Medline: 28698939]
27. Zuk G, Reinisch KB, Raptis DA, Fertsch S, Guggenheim M, Palma AF. Dupuytren disease: is there enough comprehensive patient information on the internet? Interact J Med Res 2017 Jun 22;6(1):e7 [FREE Full text] [doi: 10.2196/ijmr.7822] [Medline: 28642214]
28. Palma AF, Zuk G, Raptis DA, Franck S, Eylert G, Frueh FS, et al. Quality of information for women seeking breast augmentation in the Internet. J Plast Surg Hand Surg 2016 Oct;50(5):262-271. [doi: 10.3109/2000656X.2016.1154469] [Medline: 26985701]
29. Fox S. Health information online. Pew Internet & American Life Project. 2005 Apr 23. URL: http://www.pewinternet.org/2005/05/17/health-information-online/ [accessed 2020-11-01]

Abbreviations

BMA: British Medical Association
EQIP: Ensuring Quality Information for Patients
IPDAS: International Patient Decision Aids Standards

Edited by G Eysenbach; submitted 17.07.20; peer-reviewed by D Saleh, K Aguirre; comments to author 10.08.20; revised version received 07.09.20; accepted 17.03.21; published 26.03.21.

Please cite as:
Ghani S, Fan KS, Fan KH, Lenti L, Raptis D
Using the Ensuring Quality Information for Patients Tool to Assess Patient Information on Appendicitis Websites: Systematic Search and Evaluation
J Med Internet Res 2021;23(3):e22618
URL: https://www.jmir.org/2021/3/e22618
doi:10.2196/22618
PMID:33729160

Original Paper

Implementation of the Operating Room Black Box Research Program at the Ottawa Hospital Through Patient, Clinical, and Organizational Engagement: Case Study

Sylvain Boet[1,2,3,4], MD, PhD; Cole Etherington[1,2], PhD; Sandy Lam[1,2], RRT, MSc; Maxime Lê[5], BHSc, MA; Laurie Proulx[5], BA; Meghan Britton[6], BScN; Julie Kenna[6], BScN; Antoine Przybylak-Brouillard[1,2], MA; Jeremy Grimshaw[2], MBChB, PhD; Teodor Grantcharov[7,8], MD, PhD; Sukhbir Singh[9], MD

[1]Department of Anesthesiology and Pain Medicine, University of Ottawa, Ottawa, ON, Canada
[2]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada
[3]Department of Innovation in Medical Education, University of Ottawa, Ottawa, ON, Canada
[4]Francophone Affairs, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada
[5]Patient and Family Advisory Council, The Ottawa Hospital, Ottawa, ON, Canada
[6]Main Operating Room, The Ottawa Hospital, Ottawa, ON, Canada
[7]Department of General Surgery, University of Toronto, Toronto, ON, Canada
[8]Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON, Canada
[9]Department of Obstetrics, Gynecology, and Newborn Care, University of Ottawa, Ottawa, ON, Canada

Corresponding Author:
Sylvain Boet, MD, PhD
Department of Anesthesiology and Pain Medicine
University of Ottawa
501 Smyth Road
Ottawa, ON, K1H 8L6
Canada
Phone: 1 613 798 5555 ext 78187
Email: sboet@toh.ca

Abstract

Background: A large proportion of surgical patient harm is preventable; yet, our ability to systematically learn from these incidents and improve clinical practice remains limited. The Operating Room Black Box was developed to address the need for comprehensive assessments of clinical performance in the operating room. It captures synchronized audio, video, patient, and environmental clinical data in real time, which are subsequently analyzed by a combination of expert raters and software-based algorithms. Despite its significant potential to facilitate research and practice improvement, there are many potential implementation challenges at the institutional, clinician, and patient level. This paper summarizes our approach to implementation of the Operating Room Black Box at a large academic Canadian center.

Objective: We aimed to contribute to the development of evidence-based best practices for implementing innovative technology in the operating room for direct observation of the clinical performance by using the case of the Operating Room Black Box. Specifically, we outline the systematic approach to the Operating Room Black Box implementation undertaken at our center.

Methods: Our implementation approach included seeking support from hospital leadership; building frontline support and a team of champions among patients, nurses, anesthesiologists, and surgeons; accounting for stakeholder perceptions using theory-informed qualitative interviews; engaging patients; and documenting the implementation process, including barriers and facilitators, using the consolidated framework for implementation research.

Results: During the 12-month implementation period, we conducted 23 stakeholder engagement activities with over 200 participants. We recruited 10 clinician champions representing nursing, anesthesia, and surgery. We formally interviewed 15 patients and 17 perioperative clinicians and identified key themes to include in an information campaign run as part of the implementation process. Two patient partners were engaged and advised on communications as well as grant and protocol development. Many anticipated and unanticipated challenges were encountered at all levels. Implementation was ultimately successful, with the Operating Room Black Box installed in August 2018, and data collection beginning shortly thereafter.

Conclusions: This paper represents the first step toward evidence-guided implementation of technologies for direct observation of performance for research and quality improvement in surgery. With technology increasingly being used in health care settings, the health care community should aim to optimize implementation processes in the best interest of health care professionals and patients.

(J Med Internet Res 2021;23(3):e15443) doi:10.2196/15443

KEYWORDS

patient safety; implementation science; quality improvement; health personnel; operating rooms

Introduction

Over 50% of unintentional harm to hospitalized patients occurs in the operating room [1]. A large proportion of these incidents are preventable [1-3]. However, there has been no substantial reduction in patient safety events in recent years despite numerous advances in surgical and anesthetic practice and the proliferation of practice interventions [2,4]. Without tools that can be used to systematically observe health care provider performance in clinical practice and provide relevant and timely feedback, many interventions have been limited to the simulation environment [5-8]. Though valuable, simulation has many limitations and there is consensus that direct observation in the workplace is an extremely valuable method to assess clinical performance and to determine whether knowledge and skills transfer to practice [9].

The Operating Room Black Box was developed as a technological tool to address the need for a comprehensive understanding of clinical performance in the operating room. The Operating Room Black Box captures synchronized audio, video, patient, and environmental clinical data in real time, similar to black boxes in aviation [10]. Data captured by the Operating Room Black Box are subsequently analyzed by expert raters and software-based algorithms [10]. Accordingly, this makes it possible to study intraoperative performance without the need to be physically present in the operating room and to do so in a systematic way across a high number of surgical cases. This innovation maximizes opportunities to learn from and improve everyday practice. This leading-edge innovation also offers much needed transparency in a clinical environment that has traditionally been elusive [10,11].

When introducing tools to facilitate direct observation of clinical performance [12], particularly those that involve audio-video recording, there are many potential implementation challenges at the institutional, clinician, and patient level [12,13]. For example, there may be legal and ethical considerations, impact on workflow, and concerns about privacy, confidentiality, and evaluation [13]. Like any new health care technology, implementation may also have negative or unintended consequences if not undertaken systematically and sensitively [14]. Although much work has been done to establish best practices for implementing technology in health care, it has focused on the implementation of medical devices [15,16], electronic patient records [17-20], and simulation-based education technology [21,22]. Other studies have implemented technology for assessing performance, such as smartphone apps, and measured its impact without describing the implementation process [23]. There remains little implementation guidance regarding audio-video recording technologies designed to assess performance, especially in the operating room context.

With increasing calls for direct observation of clinical skills as part of postgraduate education and continuing professional development [24,25], it is essential to understand how tools designed for this purpose may be successfully implemented. In this paper, we outline the systematic approach to Operating Room Black Box implementation undertaken at our center. Our aim is to contribute to the development of evidence-based best practices for implementing innovative tools to directly assess performance in the operating room, using the case of the Operating Room Black Box.

Methods

Context

The Ottawa Hospital is one of Canada's largest hospitals with a total of 1202 beds across 3 campuses. At the time of the study, there were 12,003 employees at the Ottawa Hospital along with 1481 physicians and midwives, 4440 nurses, 1194 residents and fellows, and 2201 researchers. Every year, approximately 35,000 surgical cases are performed at the Ottawa Hospital.

Overview of Project Development and Implementation

The first phase of the project was a preimplementation period where the research team met with the hospital leadership and the developers of the Operating Room Black Box to determine feasibility and to establish an initial implementation plan. The research team also developed draft protocols for the first series of studies that would be conducted with Operating Room Black Box data and prepared funding applications to support implementation of the Operating Room Black Box and the planned research. The second phase of the project was the implementation period. The research team launched an information campaign to inform local stakeholders about the Operating Room Black Box and worked with the developers and local clinical managers, facility planning and support service staff, and information service technicians to install the technology.

We used a patient engagement approach to ensure that the project remained patient-centered at all times. Qualitative research was conducted throughout the project to inform and refine implementation strategies and to prospectively address challenges arising during the process.

Preimplementation Period: Understanding Individual Stakeholder Perceptions With the Theoretical Domains Framework

To identify key issues to consider in Operating Room Black Box implementation, we conducted semistructured interviews with surgical patients, perioperative clinicians, and hospital administrators. The Theoretical Domains Framework (TDF) was used to inform interview guide development and data analysis. The TDF is comprised of 14 theoretical domains derived from behavior change theories that are relevant to behavior change (eg, knowledge, beliefs about capabilities, environment/resources) [26,27]. As one of the most commonly used frameworks in implementation research, the TDF is suitable for investigating potential barriers and facilitators to a particular behavior. In our case, we explored whether stakeholders (ie, patients, clinicians, administrators) would support research using the Operating Room Black Box.

A full description of the methodology has been published elsewhere [28]. Briefly, participants were recruited across each of the 3 campuses of our center, either in-person (patients) or via email (clinicians and administrators). Interviews were conducted by 2 trained interviewers who met regularly to discuss emerging themes as sample size was determined using the concept of data saturation. Saturation was defined as conducting a minimum of 8 interviewers per stakeholder group plus an additional 3 without the emergence of any new theme. In the case of hospital administrators, saturation was defined as conducting a minimum of 5 interviews plus an additional 3, given the small number of administrators at our center. Interviews were recorded, transcribed, deidentified, and imported into a qualitative analysis software (Nvivo 11, QSR International). Direct content analysis of the interviews was conducted in duplicate by 2 independent coders using a coding strategy based on the TDF. Data units (ie, several lines of text) were coded into themes within each of the TDF domains. Belief statements were generated based on these themes in order to represent common meaning across participant responses [27]. Domain relevance (ie, whether the domain/belief should be considered during implementation) was determined based on the perceived impact of the beliefs, the presence of conflicting beliefs within a specific domain, and the relative frequency of the beliefs across participant interviews [28]. Disagreements were resolved through consensus or consultation with a third researcher. Members of the research team with expertise in using the TDF along with practicing operating room clinicians reviewed the identified themes in order to ensure credibility of the data [29].

Implementation Period: Identifying Barriers and Facilitators Across Individual and Organizational Levels With the Consolidated Framework for Implementation Research

The consolidated framework for implementation research (CFIR) [30] is a conceptual framework based on published implementation theories and reported studies. It includes 5 domains and 39 constructs that can be used “as a practical guide for systematically assessing potential barriers and facilitators in preparation for implementing an innovation” [30]. As a pragmatic model for implementation [31], the CFIR is specifically designed “to guide systematic research that supports rapid-cycle evaluation of the implementation of health care delivery interventions and produces actionable evaluation findings intended to improve implementation in a timely manner” [32].

As part of our systematic approach to implementation, we documented all aspects of the implementation process. All steps of the implementation were described and diagrams were generated to summarize each phase, including research ethics, legal and contract review, organizational parties involved, procurement, installation, data flow, consent, and information campaign strategies. This documentation provided an easy-to-follow reference point to share with stakeholders and to ensure all parties shared a common understanding of the implementation processes. We then used these documents to systematically identify themes relevant to implementation of the Operating Room Black Box across all stakeholder levels within the local context of our hospital. Barriers and facilitators were classified according to the CFIR. Direct content analysis of implementation documents was carried out by a member of the research team (SL) to identify and classify themes within the 5 CFIR domains: characteristics of the intervention, characteristics of the individuals involved, inner setting, outer setting, and process of implementation. Identified themes were then confirmed by 2 additional members of the research team (SB and CE).

Using the CFIR allowed us to prospectively develop strategies to overcome certain barriers and leverage facilitators at the systemic, organizational, and individual levels. In this way, we could address the practical needs of the stakeholders in charge of implementation of the Operating Room Black Box in our hospital as they arose.

Patient Engagement Approach

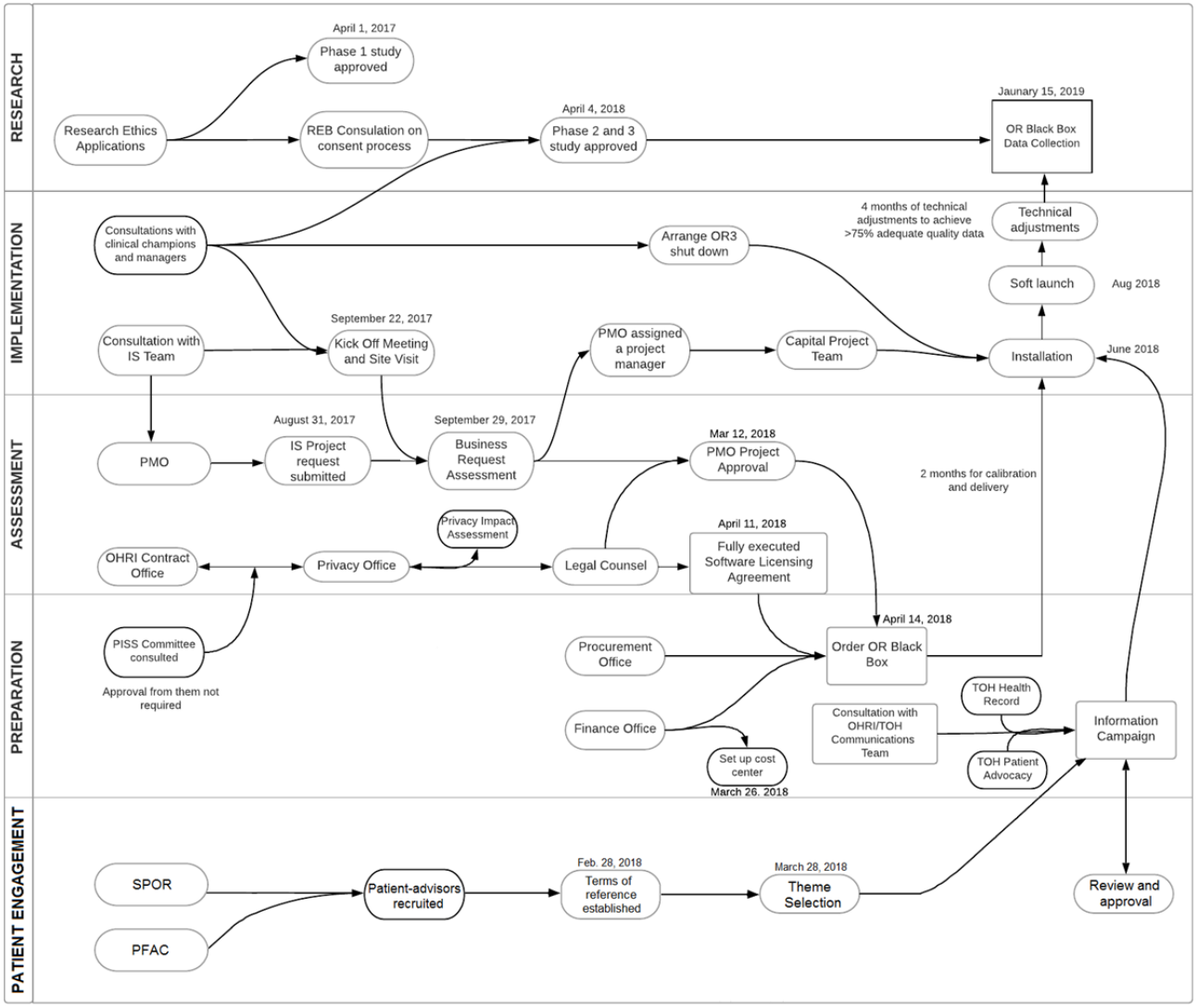
To ensure that the Operating Room Black Box implementation was patient-centered, we worked closely with the Strategy for Patient-Oriented Research (SPOR) unit located at the Ottawa Methods Centre as well as the Patient and Family Advisory Council (PFAC) in the hospital. These existing groups provided our team with resources such as relevant training materials, which allowed us to conduct patient engagement using best practices. SPOR provided us with support on grant development, patient-advisor onboarding, and patient-engagement-evaluation surveys. We elected to recruit 2 patient advisors with lived surgical experience to the research team. PFAC aided in the recruitment of patient advisors and supported logistics for our patient-engagement activities and to overcome logistical barriers (eg, food vouchers and parking passes). While these practical details may not always be considered by research teams when engaging patients, they help to facilitate sustainable long-term collaboration with patients. We planned to engage patients in all key aspects of the implementation process, from developing communication materials to reviewing grant applications and study protocols. This would set the stage for continued patient involvement as our research using the Operating Room Black Box began.

Results

Preimplementation Period

The overall implementation timeline and key activities are shown in Figure 1. The results of our experience of our implementation process are discussed below, including challenges and solutions to implementation of the technological tools studied, namely, the Operating Room Black Box.

Figure 1. Implementation process map. IS: information services; OHRI: Ottawa Hospital Research Institute; OR: operating room; PFAC: Patient and Family Advisory Council; PISS: Privacy and Information Security Steering committee; PMO: project management office; REB: research ethics board; SPOR: strategy for patient-oriented research; TDF: theoretical domains framework; TOH: The Ottawa Hospital.



Seeking Support From Hospital Leadership

In December 2015, the principal investigator (SB) met with a key senior leadership team member to discuss the Operating Room Black Box concept. Next, in January 2016, the principal investigator presented the Operating Room Black Box concept to the corporate perioperative committee, which gathered key institutional clinical leaders in surgical specialties, nursing, and anesthesiology. The committee approved the idea of an Operating Room Black Box pilot and agreed to further consider this innovation for implementation in 1 operating room.

Overview of Stakeholder Engagement Activities

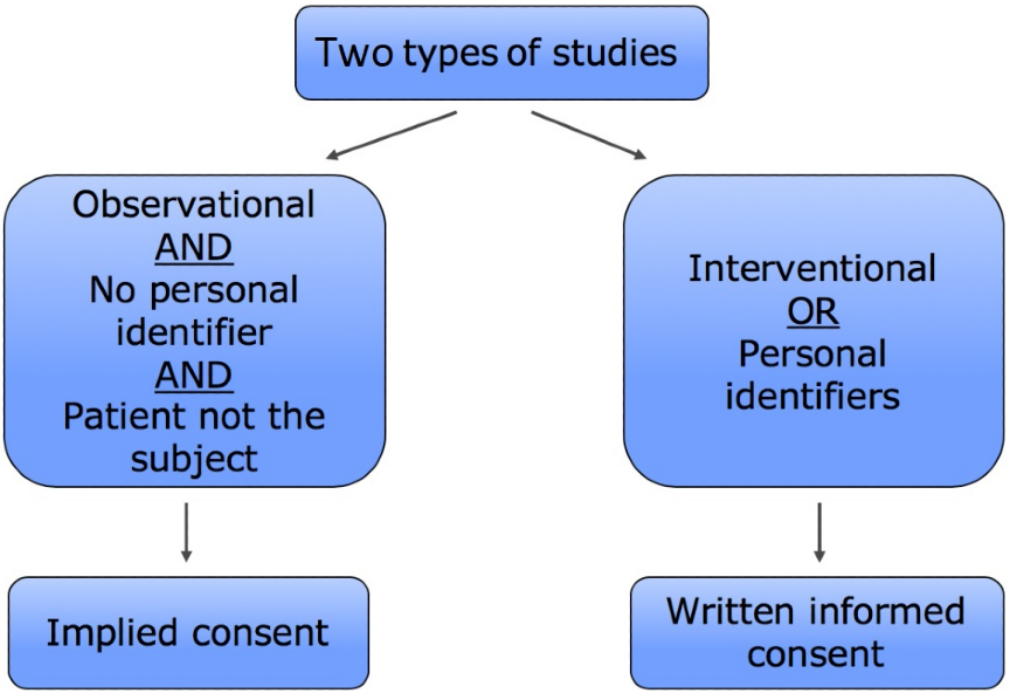
We conducted consultation meetings and presentations for various stakeholder groups in addition to dissemination activities such as an e-newsletter. These activities took place on 23 formal occasions over a 12-month period, among many additional informal meetings. Various members of the core research team were involved in each activity, which was key to connecting with different audiences. Over 200 participants have been involved in these activities, indicating the wide reach of our implementation process at our hospital. Activities involved a wide range of stakeholders—from the Research Ethics Board (REB), who were essential to determining our approach to consent, to the Health Records and Patient Advocacy departments, who would respond to any patient's request for their Operating Room Black Box recording. Multimedia Appendix 1 reports details on stakeholder engagement activities.

Privacy, Confidentiality, and Ethics

The Operating Room Black Box collects highly sensitive data both from the patient and the health care providers' perspective. Privacy, confidentiality, and consent were top priorities. To develop an optimal ethical plan, we engaged clinicians, patients,

and the Ottawa Health Science Network Research Ethics Board (OHSN-REB). The result was an approach to consent that would depend on the type of study involved (Figure 2). For observational studies that collected no personal identifiers and where the patient is not the subject of the research question, an implied consent approach is used. For interventional studies or studies that collect personal identifiers (eg, demographic information), a written informed consent is used.

Figure 2. Operating Room Black Box Research Program approach to consent.



The REB required that the implied consent approach was associated with an extensive hospital-wide information campaign. Information is first delivered to patients and clinicians prior to the surgery (through posters and pamphlets in the preoperative assessment unit, visited by patients in the weeks prior to their surgery, and institutional email). There is also a sign on the door of the room where the Operating Room Black Box is installed, and the confirmation of recording has become part of the preoperative team safety briefing. Patients and clinicians have 4 opportunities to opt out of being recorded before or within 48 hours of the surgery. It was also determined with the REB that critical situations would always be included as they represent important opportunities for learning and it would not be possible to determine who was in the room, given the high number of health care providers who would enter and leave the area, all of them wearing surgical masks. Of course, providers could still withdraw their recording up to 48 hours afterwards.

Building Frontline Support and a Team of Champions

Through informal discussions, the principal investigator—a staff anesthesiologist and researcher —shared information about the Operating Room Black Box and the research vision with colleagues in anesthesiology, nursing, surgery and perioperative clinical managers and department heads. Motivated individuals from each provider group who were supportive of the idea were invited to become research collaborators and clinician champions. We recruited 10 clinician champions in total, representing nursing, anesthesia, and surgery. These champions facilitated implementation by communicating pertinent information about the project and the technology to their colleagues and informally building support within their professions. With the assistance of our clinician champions, we also developed an internal website. The website hosted a frequently asked questions page, research information page, newsfeed, and a video featuring champions discussing the purpose and importance of the Operating Room Black Box research program.

Using a participatory and collaborative approach, our collaborators and champions contributed to the development of grant applications and study protocols via integration of their professional experience and perspectives. This approach aimed to strengthen our relationship with the collaborators while encouraging them to take ownership of the project through research engagement. As clinician champions, these individuals continued to discuss the Operating Room Black Box concept with their colleagues and were key to relaying important information to frontline staff. Champions helped to answer questions on the frontline and generate support.

Stakeholder Perceptions: TDF Interview Results

Our qualitative study of stakeholder perceptions regarding the Operating Room Black Box was approved by the OHSN-REB, Protocol #20170117-01H. This study was reported separately elsewhere [28]. In short, interviews were conducted with 15 patients across 3 hospital campuses (5 patients each), 17 perioperative clinicians, and 9 hospital administrators. Overall, patients and administrators had positive perceptions toward the Operating Room Black Box. Patients viewed the technology as a tool for their safety and as something to be expected at a teaching hospital that they trusted. Still, they expressed the importance of being provided with clear information on the

Operating Room Black Box research prior to their surgery. Administrators indicated support for the Operating Room Black Box based on its perceived fit with the institutional mandate and the expectation that using the device for research would lead to improvements in patient safety culture, processes of care, and outcomes. They also emphasized the importance of appropriate implementation and ensuring the recordings were not used punitively. Compared to patients and administrators, health care providers reported more mixed perceptions toward the Operating Room Black Box. Feelings ranged from enthusiastic support for the potential of the Operating Room Black Box to feeling skeptical or threatened. Many questions were raised about its purpose, logistical implications, and potential negative impact on the dynamics in the operating room. A desire for more information was repeatedly expressed. Clinicians were most concerned about privacy/confidentiality and possible medicolegal repercussions. Still, many reported that the Operating Room Black Box fit with their role as health care providers, aligned with institutional values, and was necessary for progress and learning. Providers also expressed their willingness to participate in the research because of their trust in the principal investigator. Based on the results of this study, we created an internal website with a "Frequently Asked Questions" page to provide more information to stakeholders, addressing their unique concerns and drawing on the perceived benefits of the Operating Room Black Box that they highlighted during the interviews.

Patient Engagement Results

Our collaboration with 2 patient advisors (ML and LP) began with the co-design of terms of reference, established between the patient advisors and the research team. The advisors were asked to assist in the development of the project and to collaborate with the research team in the development of a communications strategy. From the aforementioned interviews, the research team found 82 themes in the preliminary analysis. The researchers then presented patient advisors with a reduced number of 34 themes based on the frequency each was mentioned and perceived importance. From these, the patient advisors isolated the top 10 key themes that they determined important to convey to surgical patients and family/caregivers through communication materials that would be part of Operating Room Black Box implementation. The patient advisors collaborated in pairing patient messages with relatable graphic images and wording. This initial design session provided a framework for an informational patient poster (Multimedia Appendix 2) and pamphlet (Multimedia Appendix 3), which were then created through an iterative review process and interdisciplinary collaboration (eg, clinicians, researchers, hospital communications team, patients). Patient advisors ensured that the information and format would be relevant and easily understood by a lay audience, and their feedback influenced both the design of the materials and their placement in the care pathway (ie, made available to patients before the day of surgery). The materials created are currently available to patients in various locations at The Ottawa Hospital campuses. Of note, the design session followed an established design thinking process, which is a solution-focused methodology that employs divergent and convergent thinking of practical and creative solutions for the problems [33]. Together with the patient partners, we developed a study protocol and grant proposals for engaging patients in surgical safety research. The patient advisors provided a valuable perspective of how best to engage with patients prior to surgery and insight on communicating patient perspectives to hospital staff through newsletters and our internal website. Finally, the patient advisors engaged with the local media, the hospital newsletter, and the broader research community to share their experiences of being involved in Operating Room Black Box implementation and research projects.

Multi-Level Barriers and Facilitators: CFIR Results

Below, we report the implementation barriers and facilitators identified within each of the 5 major domains of the CFIR:

1. Characteristics of the intervention: This domain is composed of core components (the essential and indispensable elements of the intervention) and adaptable periphery (adaptable elements, structures and systems related to the intervention and the targeted organization) [31]. The lack of published evidence on the use of this intervention and its potential cost may impede certain organizations from adopting the Operating Room Black Box. At the same time, however, the Operating Room Black Box is highly adaptable and testable on a small scale in the hospital, and there is no other alternative solution to compete against it. These conditions made it favorable to implement the intervention.
2. Outer and inner settings: The outer setting refers to the economic, political, and social context surrounding an organization, whereas the inner setting refers to the structural, political, and cultural context surrounding the implementation process. The Operating Room Black Box had a favorable outer setting in that the hospital acknowledged patient safety as a top priority and has a close network with other hospitals who have adapted the same technology. In addition, the concept of the Operating Room Black Box supports the CanMEDS Physician Competency Framework [34], which is a well-recognized framework in the Canadian physician population. With regard to the inner setting, there was a favorable culture to support research as well as strong leadership engagement. The team also conducted an information campaign to create an implementation climate that facilitated access to information and boosted the receptivity of the involved individuals to the Operating Room Black Box. Barriers in the inner setting included competing with other existing projects for budget and resources and the lack of a working model between the hospital and its research institute. As a result, significant time was invested by the Research Manager to liaise between the 2 institutions. The Research Manager subsequently worked with the Vice President of Innovation and Quality at the hospital to improve the processes in the future.
3. Characteristics of the individuals involved: This domain describes the perceived control, attitudes, norms, and intentions of individuals impacted by the intervention—in this case, surgical patients and staff. An interview study was conducted prior to Operating Room Black Box

implementation to study the perceptions of these individuals. It was found that patients had positive beliefs toward the use of Operating Room Black Box to improve patient safety, and staff identified themselves with the hospital's commitment to improve patient safety and care, both of which contributed to a receptive environment during implementation. The interviews also revealed some questions and misconceptions from staff about the intervention, which were used to develop key messages for the information campaign.

4. Process of implementation: This domain describes the active change process used to promote individual and organizational use of the intervention. It is composed of 4 essential activities: planning, engaging, executing, and evaluating. The Operating Room Black Box implementation revealed several major barriers related to planning, for instance, a lack of clear administrative process to follow for the implementation of innovation in the hospital, making it challenging to develop a comprehensive implementation plan. However, the team had built strong collaborations with the hospital's capital project team and clinical departments, which greatly facilitated an effective engagement process and execution of the intervention.

The CFIR framework allowed us to assess current context through identifying major barriers and facilitators associated with implementing the Operating Room Black Box in a hospital setting. We were able to develop corresponding strategies to overcome certain barriers and play our strengths at the systemic, organizational, and individual levels. Details of barriers and facilitators using the CFIR tool are reported in Table 1.

Table 1. Barriers and facilitators to Operating Room Black Box implementation at the institutional level according to the consolidated framework for implementation research.

Domain, facilitators and barriers	Additional details
Characteristics of the Operating Room Black Box intervention	
Facilitators	
Adaptability: The platform is a highly adaptable. Its use can be tailored to local needs.	The research team secured grant funding for the purchase of the device. Long term maintenance is expected to be minimal.
Trialability: The platform is implemented on a small scale (one operating room only) and is easily reversible.	Operation of the system is simple and completely unobtrusive.
Relative advantage: There is no other existing intervention that could achieve the desired details and minimal intrusiveness offered by the platform.	N/A[a]
Barriers	
Evidence of strength and quality: New technology that lacks supporting evidence on its use to improving patient care.	Each institution has its own rules and structures related to information technology, which limited the team's ability to draw on the experiences of other centers.
Costs: There are costs associated with the purchase, installation, and maintenance of the equipment.	Before the approval of the project, there was no way for the research team to estimate costs associated with implementing the Operating Room Black Box at our institution.
Outer setting	
Facilitators	
Patient needs: Improving teamwork has been identified as a sustainable and practical way to promote patient safety.	There was a general positive environment in the outer setting that promotes the use of technology in improving patient care.
Peer pressure: The platform has been successfully implemented in 4 other hospitals in Ontario.	Evidenced by successful implementation of the Operating Room Black Box nationally and internationally.
Cosmopolitanism: Collaboration with experienced implementers to share best practices.	The lack of other alternatives to collect the same level of data in such an unobtrusive way also makes the Operating Room Black Box a favorable option.
External policy and incentives: The concept of Operating Room Black Box supports the CanMEDS Physician Competency Framework.	N/A
Inner setting	
Facilitators	
Culture: Organizational commitment to support research to improve patient care.	Letters of support were received from the Chief Executive Officer and numerous department heads to secure grant funding to purchase the device.
Readiness for implementation, leadership engagement: Overall strong support and commitment from leadership.	We established our network of support through early engagement with the senior leadership team (1 year prior to funding received).
Access to knowledge and information: A comprehensive information campaign was in place to inform affected patients and clinicians of the intervention and how it would not affect their care.	Our information campaign included emails, posters, internal website, presentations at rounds, pamphlets, excerpts in internal newsletters, stakeholder meetings, etc.
Implementation climate: The information campaign also aimed to promote positive momentum toward better practice and care through increased transparency and open discussions.	We have a structured opt-out process, which allows patients and clinicians to decline being recorded at 4 different time points. This strategy aims to increase transparency and to build a trusting relationship. This approach was developed in collaboration with clinician representatives and the Research Ethics Board.
Integration of Operating Room Black Box recording with existing work process.	
Patient advisors engaged early in the project design.	
Barriers	
Readiness for implementation, available resources: Concurrent budget cutting and other competing projects at the institutional level.	Lack of within-institution communication.

Domain, facilitators and barriers	Additional details
Networks and communications: Lack of a working model between the hospital and research institute for implementation of new technology into clinical practice.	The research institute's contract office faced many challenges related to the lack of an internal working model to collaborate with the hospital's contract office and to determine who will be leading the negotiation of the project's contract component. The Operating Room Black Box involves both research and clinical practice and therefore required approvals from both the Research Ethics Board and hospital administration. However, there was no standard procedure for the research team to follow.
Characteristics and attitudes of clinicians, patients, and senior leadership	
Facilitators	
Knowledge and beliefs about the intervention: Patients are open to the initiative.	Interviews with 15 surgical patients across the hospital's 3 campuses confirmed support and appreciation for the Operating Room Black Box.
Individual identification with organization: Shared staff commitment to improve patient safety and care.	Interviews with 17 perioperative clinicians and 9 hospital administrators identified a desire for progress and improving patient care (paper under final peer review). Patient advisors supported implementation.
Barriers	
Knowledge and beliefs about the intervention: Clinician skepticism regarding the value of new technology and perceived lack of trust in hospital management.	The interviews conducted also revealed that clinicians had many questions and misconceptions related to the use of the technology.
Operating Room Black Box implementation process	
Facilitators	
Engaging: Collaboration with the hospital's capital project team on the installation.	A peer-to-peer approach in communicating Operating Room Black Box progress was particularly useful.
Engaging: Use of an information campaign to ensure that all affected patients and clinicians are well informed.	Rather than sending out Operating Room Black Box communications through the research team, we collaborated with project champions and department leaders, who helped distribute Operating Room Black Box–related information.
Executing: Use of soft launch to stress test the data collection protocol. Standardization of communication process for anticipated patient inquiries.	We believed that people were more responsive and felt more comfortable expressing their questions or concerns to their professional peers than to the research team directly. The research team ensured that any expressed concerns were addressed and that any required opt-out paper work was filled out, hence promoting a positive environment to discuss the Operating Room Black Box with professional peers, while minimizing the extra burden on our project champions.
Planning and engaging: Kick-off meeting and regular newsletters.	N/A
Early and ongoing engagement of patient advisors	N/A
Communication strategy developed and accounted for various audiences.	We created a one-page process flow map and training materials to ensure that key actors and assessors were aware of the "big picture," and the research team filled in the gaps when questions were raised.
Barriers	
Planning: Lack of knowledge of administrative process in the hospital. Unawareness of new committees and services that need to be informed of the Operating Room Black Box	N/A
Executing: Participants are free to opt out from the program, making it impossible to predict participation rate. Implementation limited to some (not all) clinicians, creating multiple workflows for the same process. Hidden costs.	N/A

[a]N/A: not applicable.

Implementation Period

The Operating Room Black Box implementation process was initially anticipated to be 6 months but it took 15 months due to unanticipated institutional barriers. A solution had to be found to each of these barriers, drawing upon implementation facilitators related to our planned research with the Operating Room Black Box and the technology tool installation. To allow for installation, the operating room in which the Operating Room Black Box was to be installed was closed for 7 days. The

Operating Room Black Box was installed on June 29, 2018. On the door of the operating room, a sign informing all individuals that the device was present in that operating room was displayed (Multimedia Appendix 4). Following installation and testing, recording commenced in August 2018. Study data collection began in January 2019.

Discussion

This paper summarizes our approach to implementation of the Operating Room Black Box at a large academic Canadian center. This approach may be useful to guide implementation of technological tools for direct observation of clinical performance at other centers. Specifically, our approach highlights the utility of engaging stakeholders early in the implementation process and identifying barriers and facilitators across individual and organizational levels. Key strategies for successful implementation of technological tools for observing clinical performance, based on our experience with the Operating Room Black Box, are summarized in Textbox 1.

Textbox 1. Key strategies for successful implementation of the technological tools for direct observation of clinical performance.

- Engage stakeholders at the earliest phase and throughout the implementation process
- Assess all potential associated costs in advance of implementation
- Apply for installation funding
- Engage stakeholders at all levels (patients, frontline clinicians, hospital administrators, research, ethics, privacy, information technology, etc)
- Identify champions across professions
- Develop a user-centered communication and implementation plan (ie, tailored to patients and health care providers of local institution)
- Learn and map the local organization and roles and responsibilities across departments (eg, which offices need to be informed and involved)
- Include a “buffer” period in project timelines in case of unexpected delays
- Draw on implementation theory (eg, Theoretical Domains Framework, Consolidated Framework for Implementation Research)
- Develop collaborative research partnerships
- Document the implementation process
- Develop a patient engagement plan

When considering introducing technological tools for direct observation of performance in a clinical context, it is instrumental to engage with key stakeholders and funding agencies as early as possible in order to procure all necessary resources for implementation. Although we accounted for the cost of the tool itself, we did not anticipate several secondary costs associated with the project (eg, engineering time for modifications to the operating room, operating room closure during installation). Thus, it is critical to consider all potential costs prior to engaging in implementation of any technological tool in order to improve efficiency and reduce additional costs.

The partnerships we developed across stakeholder groups proved to be critical in obtaining all required approvals, finding solutions to barriers, and acquiring sufficient human and financial resources. Stakeholder engagement is a best practice that is applicable across technological tools. This is supported by the broader literature on implementing technology in health care [35] as well as guidelines in medical education for implementing methods of directly observing clinical practice [36]. In the case of the Operating Room Black Box, stakeholder engagement was central to the implementation success, particularly as challenges were encountered across multiple levels, from individual consent processes to regulatory systems.

During our implementation of the Operating Room Black Box, we simultaneously sought support from both senior leadership and frontline clinicians and made sure to listen to and incorporate feedback whenever possible. This facilitated a collaborative and solution-focused implementation approach. For example, our implied consent model emerged from consultation with the OHSN-REB, patients, clinical leaders, and frontline clinicians. Engaging frontline clinicians through a formal qualitative study and finding clinician champions also played a significant role during implementation, as concerns could be proactively addressed. Providing various opportunities for stakeholders to be heard and for stakeholders to receive information also helped to reduce any resistance that may have initially been present. Building a solid frontline support team may therefore be essential to creating positive momentum during the implementation phase of any technological tool. This type of approach could also facilitate knowledge translation and dissemination of results from the data gathered.

Beyond our collaboration with perioperative clinicians of all professions, administrators, and researchers, a second notable strength of our implementation approach was patient engagement. While patient engagement in other areas of health care has been well-documented [37], there is a clear knowledge gap regarding patient engagement in surgical patient safety research. This is problematic, in particular, given that patients are most vulnerable and unable to advocate for their own needs while under anesthesia. Our partnership with our patient advisors was essential for making our information campaign patient-centered and for aligning research priorities with those of patients. Significant efforts were made to plan for patient engagement activities, including detailing roles and responsibilities, timelines, level of effort, contributions, outreach, and evaluation. These evaluations allowed regular

assessments to occur throughout the project and indicated both a positive experience by all team members and a meaningful impact on the project. Patient advisors also attended various local and international conferences, which helped to build momentum across institutions that are using or planning to use the Operating Room Black Box, fostering future collaborations to engage patients in surgical safety research on a larger scale.

Using technology for direct observation of clinical performance has many potential benefits for both patients and health care providers [11,38]. For example, data from the Operating Room Black Box system can be used to enhance training by learning from safety threats and resiliency supports. This, in turn, may improve processes of care and patient outcome. Of course, there are legal aspects of video and audio capture in health care, which warrant discussion. Although health care providers may have concerns about medicolegal risks, consultations with legal experts suggest recorded data would be protected under the *Healthcare Quality Improvement Act, 1986* (HCQIA) in the United States and the *Quality of Care Information Protection Act, 2016* (QCIPA) in Canada. The HCQIA protects medical professionals from prosecution for conduct that undergoes peer review, while QCIPA allows quality improvement matters, including critical incidents, to be openly discussed among health professionals. Essentially, each of these pieces of legislation aim to further improve the quality through open communication without fear of reprisal. As with any new form of data collection, time is needed to continue to study and address its implications. That being said, as long as technology such as the Operating Room Black Box continues to be implemented by and for frontline health care providers in partnership with patients, it is likely to result in innovative quality improvement that simultaneously protects operating room teams, hospitals, and patients—all working together toward the same common goal of quality of care.

We recognize that the approach described in this paper represents the implementation experience of a unique technological tool at 1 large academic Canadian center only and that there may be center-specific factors, which may be important to explore prior to implementation. Nonetheless, the overall general approach and documentation process that we report in this paper may be useful to other centers aiming to implement technological tools for direct observation of clinical performance in the operating room.

In conclusion, this paper represents the first step toward evidence-guided implementation of technologies for direct observation of performance for research and quality improvement in surgery. With technology increasingly being used in health care settings, the health care community should aim to optimize implementation processes in the best interest of health care professionals and patients.

Acknowledgments

The authors would like to thank all of the hospital stakeholders for their support of Operating Room Black Box initiatives.

Conflicts of Interest

TG is the founder and equity holder of Surgical Safety Technologies Inc.

Multimedia Appendix 1
Operating Room Black Box implementation: consultations and presentations.
[DOCX File , 16 KB - jmir_v23i3e15443_app1.docx]

Multimedia Appendix 2
Informational poster for patients.
[PNG File , 335 KB - jmir_v23i3e15443_app2.png]

Multimedia Appendix 3
Informational pamphlet for patients.
[PNG File , 431 KB - jmir_v23i3e15443_app3.png]

Multimedia Appendix 4
Poster on the operating room door.
[PNG File , 33 KB - jmir_v23i3e15443_app4.png]

References

1. Zegers M, de Bruijne MC, de Keizer B, Merten H, Groenewegen PP, van der Wal G, et al. The incidence, root-causes, and outcomes of adverse events in surgical units: implication for potential prevention strategies. Patient Saf Surg 2011 May 20;5:13 [FREE Full text] [doi: 10.1186/1754-9493-5-13] [Medline: 21599915]
2. Measuring patient harm in Canadian hospitals. Canadian Institute for Health Information and Canadian Patient Safety Institute. 2016. URL: https://www.cihi.ca/sites/default/files/document/hospital_harm_technical_notes_en.pdf [accessed 2019-03-07]

3. Hu Y, Arriaga AF, Peyre SE, Corso KA, Roth EM, Greenberg CC. Deconstructing intraoperative communication failures. J Surg Res 2012 Sep;177(1):37-42 [FREE Full text] [doi: 10.1016/j.jss.2012.04.029] [Medline: 22591922]
4. Kapur N, Parand A, Soukup T, Reader T, Sevdalis N. Aviation and healthcare: a comparative review with implications for patient safety. JRSM Open 2016 Jan;7(1):2054270415616548 [FREE Full text] [doi: 10.1177/2054270415616548] [Medline: 26770817]
5. Salas E, Frush K. Improving Patient Safety Through Teamwork and Team Training. Oxford, UK: Oxford University Press; 2012.
6. Shapiro MJ, Morey JC, Small SD, Langford V, Kaylor CJ, Jagminas L, et al. Simulation based teamwork training for emergency department staff: does it improve clinical team performance when added to an existing didactic teamwork curriculum? Qual Saf Health Care 2004 Dec;13(6):417-421 [FREE Full text] [doi: 10.1136/qhc.13.6.417] [Medline: 15576702]
7. Phipps MG, Lindquist DG, McConaughey E, O'Brien JA, Raker CA, Paglia MJ. Outcomes from a labor and delivery team training program with simulation component. Am J Obstet Gynecol 2012 Jan;206(1):3-9. [doi: 10.1016/j.ajog.2011.06.046] [Medline: 21840493]
8. Guise J, Deering SH, Kanki BG, Osterweil P, Li H, Mori M, et al. Validation of a tool to measure and promote clinical teamwork. Simul Healthc 2008;3(4):217-223. [doi: 10.1097/SIH.0b013e31816fdd0a] [Medline: 19088666]
9. Weller JM, Bloch M, Young S, Maze M, Oyesola S, Wyner J, et al. Evaluation of high fidelity patient simulator in assessment of performance of anaesthetists. Br J Anaesth 2003 Jan;90(1):43-47 [FREE Full text] [Medline: 12488377]
10. Goldenberg MG, Jung J, Grantcharov TP. Using Data to Enhance Performance and Improve Quality and Safety in Surgery. JAMA Surg 2017 Oct 01;152(10):972-973. [doi: 10.1001/jamasurg.2017.2888] [Medline: 28813552]
11. Langerman A, Grantcharov TP. Are We Ready for Our Close-up?: Why and How We Must Embrace Video in the OR. Ann Surg 2017 Dec;266(6):934-936. [doi: 10.1097/SLA.0000000000002232] [Medline: 28338514]
12. Cheung WJ, Patey AM, Frank JR, Mackay M, Boet S. Barriers and Enablers to Direct Observation of Trainees' Clinical Performance: A Qualitative Study Using the Theoretical Domains Framework. Acad Med 2019 Jan;94(1):101-114. [doi: 10.1097/ACM.0000000000002396] [Medline: 30095454]
13. Asan O, Montague E. Using video-based observation research methods in primary care health encounters to evaluate complex interactions. Inform Prim Care 2014;21(4):161-170 [FREE Full text] [doi: 10.14236/jhi.v21i4.72] [Medline: 25479346]
14. Nielsen CP, Funch TM, Kristensen FB. Health technology assessment: research trends and future priorities in Europe. J Health Serv Res Policy 2011 Jul;16 Suppl 2:6-15. [doi: 10.1258/jhsrp.2011.011050] [Medline: 21737525]
15. Edmondson AC, Bohmer RM, Pisano GP. Disrupted Routines: Team Learning and New Technology Implementation in Hospitals. Administrative Science Quarterly 2001 Dec;46(4):685. [doi: 10.2307/3094828]
16. Sachdeva AK, Russell TR. Safe introduction of new procedures and emerging technologies in surgery: education, credentialing, and privileging. Surg Clin North Am 2007 Aug;87(4):853-66, vi. [doi: 10.1016/j.suc.2007.06.006] [Medline: 17888784]
17. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful implementation and adoption of large-scale health information technology. J Am Med Inform Assoc 2013 Jun;20(e1):e9-e13 [FREE Full text] [doi: 10.1136/amiajnl-2013-001684] [Medline: 23599226]
18. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752 [FREE Full text] [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]
19. Ludwick DA, Doucette J. Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. Int J Med Inform 2009 Jan;78(1):22-31. [doi: 10.1016/j.ijmedinf.2008.06.005] [Medline: 18644745]
20. Meyfroidt G. How to implement information technology in the operating room and the intensive care unit. Best Pract Res Clin Anaesthesiol 2009 Mar;23(1):1-14. [doi: 10.1016/j.bpa.2008.07.004] [Medline: 19449612]
21. Smith CD. Simulation Technology: A Strategy for Implementation in Surgical Education and Certification. Presence: Teleoperators & Virtual Environments 2000 Dec;9(6):632-637. [doi: 10.1162/105474600300040420]
22. Schout BMA, Hendrikx AJM, Scheele F, Bemelmans BLH, Scherpbier AJJA. Validation and implementation of surgical simulators: a critical review of present, past, and future. Surg Endosc 2010 Mar;24(3):536-546 [FREE Full text] [doi: 10.1007/s00464-009-0634-9] [Medline: 19633886]
23. Fryer JP, Teitelbaum EN, George BC, Schuller MC, Meyerson SL, Theodorou CM, et al. Effect of Ongoing Assessment of Resident Operative Autonomy on the Operating Room Environment. J Surg Educ 2018;75(2):333-343. [doi: 10.1016/j.jsurg.2016.11.018] [Medline: 28363675]
24. Massie J, Ali JM. Workplace-based assessment: a review of user perceptions and strategies to address the identified shortcomings. Adv Health Sci Educ Theory Pract 2016 May;21(2):455-473. [doi: 10.1007/s10459-015-9614-0] [Medline: 26003590]
25. Holmboe ES. Realizing the promise of competency-based medical education. Acad Med 2015 Apr;90(4):411-413. [doi: 10.1097/ACM.0000000000000515] [Medline: 25295967]

26. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
27. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A. Making psychological theory useful for implementing evidence based practice: a consensus approach. Qual Saf Health Care 2005 Feb;14(1):26-33 [FREE Full text] [doi: 10.1136/qshc.2004.011155] [Medline: 15692000]
28. Etherington C, Usama A, Patey AM, Trudel C, Przybylak-Brouillard A, Presseau J, et al. Exploring stakeholder perceptions around implementation of the Operating Room Black Box for patient safety research: a qualitative study using the theoretical domains framework. BMJ Open Qual 2019;8(3):e000686 [FREE Full text] [doi: 10.1136/bmjoq-2019-000686] [Medline: 31428707]
29. Shenton AK. Strategies for ensuring trustworthiness in qualitative research projects. EFI 2004 Jul 19;22(2):63-75. [doi: 10.3233/EFI-2004-22201]
30. Technical assistance for users of the CFIR framework. Consolidated Framework for Implementation Research. 2018. URL: https://cfirguide.org/ [accessed 2019-03-07]
31. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
32. Keith RE, Crosson JC, O'Malley AS, Cromp D, Taylor EF. Using the Consolidated Framework for Implementation Research (CFIR) to produce actionable findings: a rapid-cycle evaluation approach to improving implementation. Implement Sci 2017 Feb 10;12(1):15 [FREE Full text] [doi: 10.1186/s13012-017-0550-7] [Medline: 28187747]
33. Design thinking: A method for creative problem solving. IDEOU. URL: https://www.ideou.com/pages/design-thinking [accessed 2019-03-07]
34. Frank L, Snell J, Sherbino J. CanMEDS 2015 physician competency framework. 2015. URL: http://canmeds.royalcollege.ca/uploads/en/framework/CanMEDS%202015%20Framework_EN_Reduced.pdf [accessed 2019-07-02]
35. Acharya S, Werts N. Toward the Design of an Engagement Tool for Effective Electronic Health Record Adoption. Perspect Health Inf Manag 2019;16(Winter):1g [FREE Full text] [Medline: 30766458]
36. Kogan JR, Hatala R, Hauer KE, Holmboe E. Guidelines: The do's, don'ts and don't knows of direct observation of clinical skills in medical education. Perspect Med Educ 2017 Oct;6(5):286-305 [FREE Full text] [doi: 10.1007/s40037-017-0376-7] [Medline: 28956293]
37. Prey JE, Woollen J, Wilcox L, Sackeim AD, Hripcsak G, Bakken S, et al. Patient engagement in the inpatient setting: a systematic review. J Am Med Inform Assoc 2014;21(4):742-750 [FREE Full text] [doi: 10.1136/amiajnl-2013-002141] [Medline: 24272163]
38. Catchpole K, Neyens DM, Abernathy J, Allison D, Joseph A, Reeves ST. Framework for direct observation of performance and safety in healthcare. BMJ Qual Saf 2017 Dec;26(12):1015-1021 [FREE Full text] [doi: 10.1136/bmjqs-2016-006407] [Medline: 28971880]

Abbreviations

CFIR: consolidated framework for implementation research
HCQIA: Healthcare Quality Improvement Act, 1986
OHSN-REB: Ottawa Health Science Network Research Ethics Board
PFAC: Patient and Family Advisory Council
QCPIA: Quality of Care Information Protection Act, 2016
REB: research ethics board
SPOR: strategy for patient-oriented research
TDF: theoretical domains framework

Edited by G Eysenbach; submitted 11.07.19; peer-reviewed by A Joseph, N Zakaria, D John; comments to author 21.10.19; revised version received 11.12.19; accepted 26.02.20; published 16.03.21.

Please cite as:
Boet S, Etherington C, Lam S, Lê M, Proulx L, Britton M, Kenna J, Przybylak-Brouillard A, Grimshaw J, Grantcharov T, Singh S Implementation of the Operating Room Black Box Research Program at the Ottawa Hospital Through Patient, Clinical, and Organizational Engagement: Case Study
J Med Internet Res 2021;23(3):e15443
URL: https://www.jmir.org/2021/3/e15443
doi:10.2196/15443
PMID:33724199

Original Paper

e-Mental Health Program Usage Patterns in Randomized Controlled Trials and in the General Public to Inform External Validity Considerations: Sample Groupings Using Cluster Analyses

Samineh Sanatkar[1,2], PhD; Peter Baldwin[1], MPsychol, PhD; Kit Huckvale[1,2], PhD; Helen Christensen[1], PhD; Samuel Harvey[1], MBBS, MRCGP, MRCPsych, FRANZCP, PhD

[1]Black Dog Institute, The University of New South Wales Sydney, Randwick, Australia

[2]School of Psychiatry, The University of New South Wales Sydney, Randwick, Australia

Corresponding Author:
Samineh Sanatkar, PhD
Black Dog Institute
The University of New South Wales Sydney
Hospital Road
Randwick, 2031
Australia
Phone: 61 9382 4368
Email: s.sanatkar@unsw.edu.au

Abstract

Background: Randomized controlled trials (RCTs) with vigorous study designs are vital for determining the efficacy of treatments. Despite the high internal validity attributed to RCTs, external validity concerns limit the generalizability of results to the general population. Bias can be introduced, for example, when study participants who self-select into a trial are more motivated to comply with study conditions than are other individuals. These external validity considerations extend to e-mental health (eMH) research, especially when eMH tools are designed for public access and provide minimal or no supervision.

Objective: Clustering techniques were employed to identify engagement profiles of RCT participants and community users of a self-guided eMH program. This exploratory approach inspected actual, not theorized, RCT participant and community user engagement patterns. Both samples had access to the eMH program over the same time period and received identical usage recommendations on the eMH program website. The aim of this study is to help gauge expectations of similarities and differences in usage behaviors of an eMH tool across evaluation and naturalistic contexts.

Methods: Australian adults signed up to myCompass, a self-guided online treatment program created to reduce mild to moderate symptoms of negative emotions. They did so either by being part of an RCT onboarding (160/231, 69.6% female) or by accessing the program freely on the internet (5563/8391, 66.30% female) between October 2011 and October 2012. During registration, RCT participants and community users provided basic demographic information. Usage metrics (number of logins, trackings, and learning activities) were recorded by the system.

Results: Samples at sign-up differed significantly in age (P=.003), with community users being on average 3 years older (mean 41.78, SD 13.64) than RCT participants (mean 38.79, SD 10.73). Furthermore, frequency of program use was higher for RCT participants on all usage metrics compared to community users through the first 49 days after registration (all P values <.001). Two-step cluster analyses revealed 3 user groups in the RCT sample (Nonstarters, 10-Timers, and 30+-Timers) and 2 user groups in the community samples (2-Timers and 20-Timers). Groups seemed comparable in patterns of use but differed in magnitude, with RCT participant usage groups showing more frequent engagement than community usage groups. Only the high-usage group among RCT participants approached myCompass usage recommendations.

Conclusions: Findings suggested that external validity concerns of RCT designs may arise with regards to the predicted magnitude of eMH program use rather than overall usage styles. Following up RCT nonstarters may help provide unique insights into why individuals choose not to engage with an eMH program despite generally being willing to participate in an eMH evaluation study. Overestimating frequency of engagement with eMH tools may have theoretical implications and potentially impact economic considerations for plans to disseminate these tools to the general public.

(J Med Internet Res 2021;23(3):e18348) doi:10.2196/18348

KEYWORDS

e-mental health; engagement patterns; external validity; randomized controlled trial; community sample

Introduction

Well-designed randomized controlled trials (RCTs) are widely seen as the gold standard for determining treatment efficacy as random assignment to either a treatment or control group allows for the isolation of the treatment effect from both known and unknown confounding factors [1]. Although the importance of RCTs in establishing internal validity is undisputed, researchers have pointed out the external validity concerns of RCTs [2]. These concerns relate to participant selection, attention, retention, researcher contact, and specifics and frequency of data collection—all of which can limit the generalizability of findings to the general public.

Such external validity considerations may be even more justified when considering those RCTs that evaluate e-mental health (eMH) programs, which are designed to deliver effective, scalable mental health care in the community [3-6]. For example, RCTs of eMH programs often selectively recruit from online communities [7] and provide a level of direction for adhering to usage recommendations (eg, reminders) that users in the general public will not encounter, particularly in self-guided eMH programs. These top-down practices affecting user attrition have been described as "push factors" [8]. RCT participants may also be particularly motivated to follow the research protocol and have researcher contact, which likely differs from how the general public experience an eMH program. Furthermore, deterministic approaches to evaluating eMH program effectiveness do not necessarily reflect the ever-changing eMH landscape [5]. As Sieverink and colleagues [9] pointed out, RCTs lack the ability to inform about processes (ie, how behavior evolves between the pre-, post-, and follow-up assessments) or which program components (or combination thereof) contribute to improvements in mental health outcomes. It is therefore pivotal for eMH interventions to show that the treatment effects initially shown in RCTs can translate into real-world benefits in practicable ways.

Despite the need for eMH programs to be applicable to real-world conditions, information on the external validity of eMH treatment outcomes is not widely available. In a recent systematic review of digital interventions addressing comorbid depressive symptoms and substance use, only 1 of the 6 studies examined reported on the comparability of the sample used to the wider population [10]. Similarly, in a review paper of mobile apps promoting physical activity, Blackman and colleagues [11] found that all mobile health intervention studies considered (N=20) reported on treatment effectiveness, but only 4 of these studies reported on how representative the study sample was for the target population.

This problem extends to eMH program design. A considerable body of research examining the effects of different eMH design features on behavioral changes has not addressed the question of how applicable these findings are once the programs are disseminated [3]. Although at least some studies address limitations to the generalization of eMH trial findings, studies examining how or if study protocols influence eMH engagement behaviors are rare. Arguably, whether or not a study protocol influences participant behavior constitutes another important factor in establishing the external validity of eMH findings [4]. One comparison between RCT and real-world uptake was undertaken with moodgym, an eMH program aimed at reducing anxiety and depression. It showed that public registrants were less likely than RCT participants to complete the recommended number of treatment modules [12]. Interestingly, however, symptom reductions over time were comparable between both RCT and community user groups, raising the question of whether a reduced protocol may yield similar benefits.

This short paper attempts to deepen the discussion on the influence of the RCT environment on eMH engagement behavior by presenting engagement patterns of RCT participants and community users of an eMH tool, called myCompass, side by side. Our aim is to explore whether patterns of program engagement differ between RCT participants who receive usage recommendations as part of being involved in an evaluation study versus users in the general community who receive the same usage recommendations only on the myCompass homepage. The goal of this paper is to examine engagement rather than outcomes; therefore, we are presenting cluster analysis findings that help visualize behavioral patterns rather than quantify differences in health and well-being.

Methods

Program Description

The present analyses are based on the first version of myCompass, an eMH program that was available to all Australians between 2011 and 2018. myCompass is hosted and run by the Black Dog Institute, and funded by the Australian Department of Health. Version 1 of this self-guided program offered online mental health resources and activities to address mild to moderate symptoms of depression, anxiety, and stress. Core functionalities of myCompass were the daily tracking and learning activities components. The tracking function allowed users to track up to 3 moods, behaviors, or cognitions (eg, sadness, alcohol consumption, worry) in real time. Users indicated their current states on an interval scale from low (0) to high (10). Another main function of myCompass was learning activities. Learning activities were a set of 14 modules which aimed at aiding beneficial behaviors such as goal setting, sleep quality, or managing fear and anxiety. Each module was split into 2 to 3 sessions to promote skill-building exercises over the course of several days and took about 10 to 15 minutes to complete online.

Samples

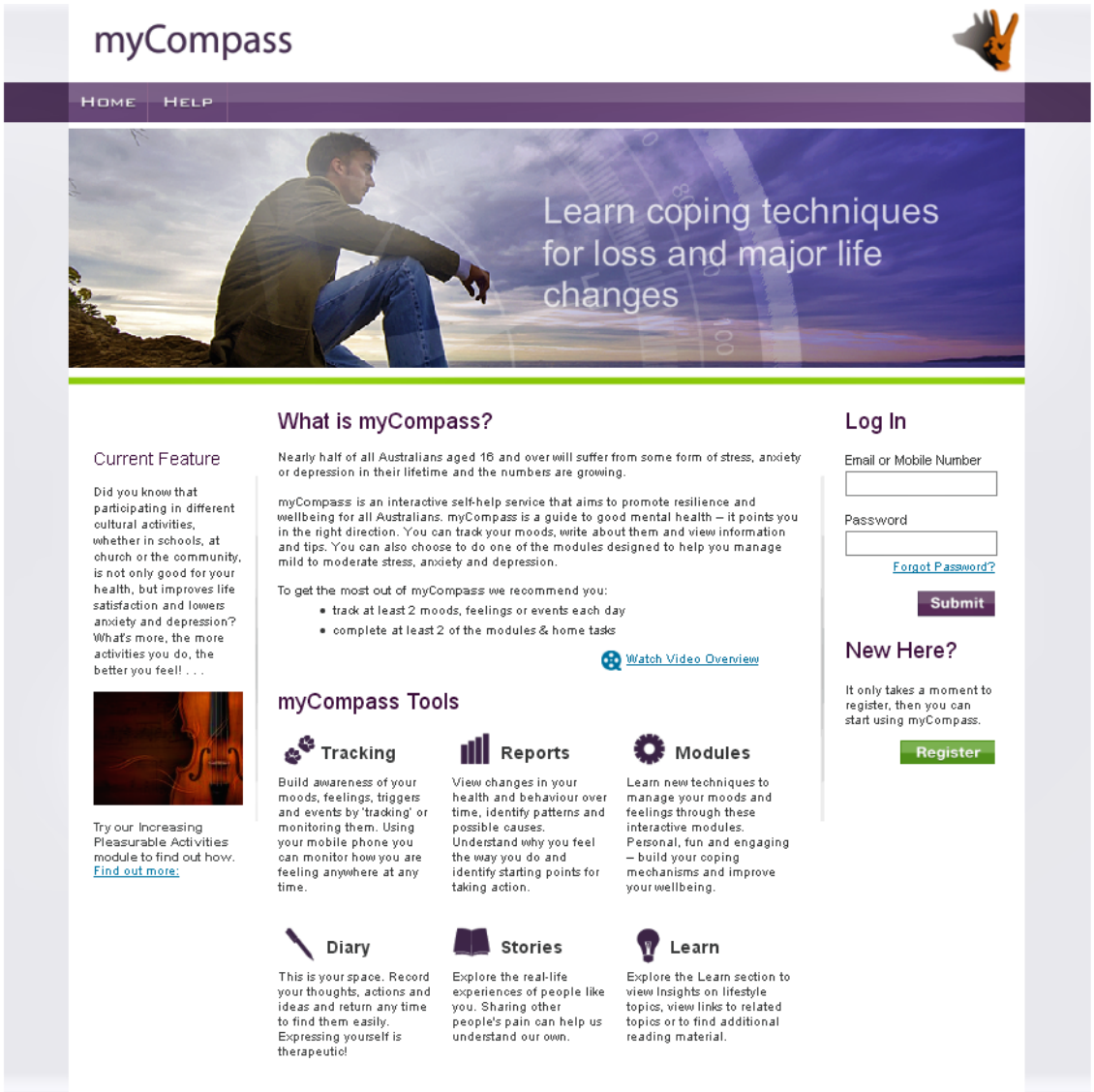
Our study considered engagement data from 2 distinct samples: (1) an RCT participant sample and (2) a naturalistic community sample of general public users who freely adopted myCompass. All participants and users registered to version 1.0 of myCompass between October 2011 and October 2012. Trial

participants and community users freely registered either to the research study or directly to the myCompass website. Trial recruitment took place through multiple avenues, such as social media posts on Facebook and announcements on the Black Dog Institute volunteer research register [13], while the myCompass website could be found spontaneously through internet search engines. RCT participants and general public users received the same usage recommendations on the myCompass home page of completing at least 2 modules and using the tracking function once a day (see Figure 1). All users, independent of whether or not they were part of the research study, set the frequency of program use reminders in line with their own preferences on the myCompass website. However, only research participants were able contact research staff.

The RCT sample comprised 231 participants allocated to the eMH treatment group for the initial evaluation study of myCompass [13]. Exclusion criteria for entering the study were being younger than 18 or older than 75 years, not possessing an internet-enabled mobile phone, not having access to a computer with internet and email, and showing either minimal or severe symptoms of depression and anxiety as determined by symptom scores on the Depression Anxiety and Stress Scale [14].

The community sample comprised 8391 adults who registered for myCompass on their own accord. Community members needed to provide a valid email address and mobile phone number to verify their willingness to register to the program. As part of the profile setup within the myCompass program, RCT participants and community users alike completed assessments of common mental health symptoms, which formed the basis for tracking recommendations made by the program [13]. If scores indicated severe distress (ie, depression, anxiety, or a stress scores of 8 or higher out of 10) or suicidal ideation, the sign-up process was terminated and individuals were redirected to the Black Dog Institute website with information on how to seek immediate support.

Figure 1. Screenshot of myCompass homepage with usage recommendations.



Measurement of User Engagement

We selected 4 usage metrics to assess user engagement. These were number of logins, number of logged trackings (irrespective of how many items were tracked at any one time), number of learning activities started, and number of learning activities completed.

Statistical Analysis

All analyses were conducted in SPSS version 25 (IBM Corp). Participants using myCompass as part of the RCT and those in the community sample were initially compared using chi-square and *t* tests. We then conducted 2 two-step cluster analyses to determine distinct usage groups within the RCT participant and community user samples. A two-step cluster analysis allows for the detection of naturalistic (ie, not hypothesis-driven) groupings within a data set by way of examining distances between data points (step 1). Based on these distance calculations, an algorithm (step 2) determines the number of groupings or clusters for each data set. In the current analysis, we selected the log-likelihood distance measure and the Schwarz Bayesian information criterion statistic to determine the most appropriate number of clusters.

Results

Demographic Information

Data were available for 8622 users of myCompass. Table 1 shows the basic demographic information and online engagement behaviors of individuals taking part in the myCompass RCT and those signing up to myCompass via the program's website. Although both groups had similar gender ratios (RCT: 160/231, 69.6% female; general public: 5563/8391, 66.30% female; *P*=.30), community users were on average 3 years older than RCT participants (*P*=.003). Notably, all average access and engagements statistics were higher for RCT participants than for general public users (all *P* values <.001).

Table 1. Mean, SD, and between-group statistics on demographic information and usage behavior through 49 days (N=8622).

Variable	RCT[a] sample (n=231), mean (SD)	Community sample (n=8391), mean (SD)	*F* test	*P* values	d
Age (years)	38.79 (10.73)	41.78 (13.64)	9.07	.003	0.20
Female (%)	1.70 (0.46)	1.66 (0.47)	1.07[b]	.30	N/A
49-day logins (n)	11.26 (15.78)	3.90 (6.91)	229.21	<.001	1.01
49-day trackings (n)	11.26 (10.42)	2.57 (5.95)	343.48	<.001	1.24
49-day mod[c] started (n)	1.14 (1.62)	0.53 (0.85)	113.81	<.001	0.71
49-day mod completed (n)	1.02 (1.53)	0.12 (0.53)	555.85	<.001	1.57

[a]RCT: randomized controlled trial.

[b]Chi-square statistic.

[c]mod: modules.

Two-Step Cluster Analyses

A number of RCT participants allocated to the myCompass group did not proceed to register to myCompass (73/231, 31.6%) and therefore did not provide any engagement data. To make the RCT group more comparable to the general public sample (where each person registered to myCompass and provided engagement data), we removed these participants from the sample used for the cluster analysis. Table 2 shows the results of the cluster analysis among RCT participants who were allocated to myCompass for the duration of the 7-week intervention period. The cluster analysis yielded 2 distinct user groups among the RCT participants. Results of a multivariate analysis of variance confirmed significant cluster group differences on all 4 usage variables (all *P* values <.001; see Table 2), indicating that the clustering procedure was successful in establishing distinct user groups.

The first and larger user group (114/158, 72.2%) were "10-Timers" who logged into their assigned program around 9 times (mean 8.84, SE 2.01) and used myCompass mainly for tracking (mean 7.77, SE 1.83) rather than completing learning activities (mean 0.23, SE 0.21). The second user group (44/158, 27.8%) were "30+-Timers", who logged into myCompass every other day (mean 36.20, SE 1.71). When the 30+-Timers logged on, they used the program's tracking function (mean 34.59, SE 1.56) in addition to starting and completing around three learning activities (modules started: mean 3.36, SE 0.19; modules completed: mean 2.67, SE 0.18) over the course of 7 weeks.

Table 2. Cluster analysis groupings of 158 myCompass randomized controlled trial participants' usage behaviors and multivariate analysis of variance testing of the significant difference between clusters.

Variables	Nonstarters (n=73), mean (SE)	10-Timers (n=114), mean (SE)	30+-Timers (n=44), mean (SE)	*F* test (1, 156)	*P* values
Age (years)	38.79 (10.73)[a]	37.08 (10.71)	41.40 (8.09)	5.82	.02
Female (%)	1.70 (0.46)[a]	1.70 (0.46)	1.66 (0.48)	0.24	.62
Logins (n)	N/A	8.84 (2.01)	36.20 (1.71)	184.53	<.001
Trackings (n)	N/A	7.77 (1.83)	34.59 (1.56)	214.73	<.001
Mod[b] started (n)	N/A	0.77 (0.22)	3.36 (0.19)	145.22	<.001
Mod completed (n)	N/A	0.23 (0.21)	2.67 (0.18)	156.73	<.001

[a]Demographic information on Nonstarters is included for informational purposes and was not part of the analysis.

[b]mod: modules.

Table 3 presents the results of the cluster analysis in the general public. Similar to findings among RCT participants, the community sample cluster analysis yielded 2 distinct clusters that differed across all usage variables (all *P* values <.001). The vast majority of general public users (7681/8391, 91.54%) were "2-Timers" or individuals who entered myCompass approximately twice (mean 2.25, SE 0.17) and used the tracking function once (mean 1.13, SE 0.14). The average module completion neared zero in this group (mean 0.03, SE 0.02).

The second and considerably smaller group were the "20-Timers" (710/8391, 8.46%) who used myCompass consistently. Members of this group logged in on average 22 times (mean 21.82, SE 0.16) and used the tracking function the majority of the time (mean 18.06, SE 0.14) over a 7-week period. In addition, 20-Timers started about 2 modules (mean 1.98, SE 0.03) and completed 1 (mean 1.07, SE 0.02) during this time.

Graphical representations of the cluster solutions can be found in Figures S1 and S2 in Multimedia Appendix 1.

Table 3. Cluster analysis groupings of 8391 myCompass general public users usage behaviors and multivariate analysis of variance testing of the significant difference between clusters.

Variables	2-Timers (n=7681), mean (SE)	20-Timers (n=710), mean (SE)	*F* test (1, 8389)	*P* values
Age (years)	41.70 (13.67)	42.83 (13.33)	4.59	.03
Female (%)	1.66 (0.47)	1.67 (0.47)	0.08	.78
Logins (n)	2.25 (0.17)	21.82 (0.16)	13803.55	<.001
Trackings (n)	1.13 (0.14)	18.06 (0.14)	14057.60	<.001
Modules started (n)	0.40 (0.03)	1.98 (0.03)	3145.51	<.001
Modules completed (n)	0.03 (0.02)	1.07 (0.02)	3670.39	<.001

Discussion

This paper presents findings on individual usage behaviors for an Australian eMH tool, either as part of an RCT or as an open-access tool freely adopted by the general public. Exploratory findings reveal that the same number of usage groups emerged in both data sets: a large "lower-intensity" usage group and a smaller "higher-intensity" usage group. However, our findings revealed interesting differences between the 2 data sets that warrant consideration. First, the general community group tended to be older than the RCT group and overall used the eMH program significantly less frequently based on all usage metrics. Of further note, a considerable number of individuals registered to myCompass directly and were not part of the research trial. This could be because participation in a research study is more time intensive, as research volunteers are required to complete psychometric measures in addition to using the eMH tool. It is also possible that only a relatively smaller number of individuals were exposed to the research trial recruitment calls, while a greater number of interested individuals were able to discover the myCompass website using internet search engines.

Second, although the cluster analytical findings revealed 2 behavioral groups across both samples, the magnitude of usage was higher in the RCT sample for both usage groups. For example, the low-usage group in the RCT logged in an average of 9 times, while the low-usage group in the general community logged in only about twice on average. The high-usage group in the RCT sample was, again, not only higher in magnitude (about 35 logins on average as opposed to about 20 logins in the general community), but also proportionally bigger than that in the community sample. Specifically, 27.8% (44/158) of RCT users were identified as frequent users (30+-Timers), whereas only 8.46% (710/8391) of general public users were identified as such (20-Timers). Accordingly, the 30+-Timers RCT group completed the recommended 2 or more learning activities and came closest to the tracking recommendations of 49 logged trackings, whereas the 20-Timers community sample group clearly did not meet the tracking or learning activity recommendations. Thus, only the high-usage RCT group could

be described as "adhering" to the learning activity recommendations, and no group adhered to the tracking recommendations.

Our findings add weight to Cavanagh's [4] concerns about external validity in eMH trials, suggesting that inferences about real-world engagement from RCT data indeed should be made with caution. Although sample composition and usage patterns were comparable between the RCT and general community users, generalizing from the RCT sample would have overestimated the magnitude of real-world program engagement. One potential reason for this could be the differing motivation for eMH adoption. In our study, RCT participants seemed to be more motivated to use the eMH program and to use the core functionalities more consistently than were users in the general public. It is possible that, beyond the willingness to participate in mental health research, the aims stated in the Participant Information Statement inadvertently attracted individuals who were interested in the topic of eMH and therefore more motivated to engage with an eMH tool in general and with the activities recommended to them in particular. On the other side of the behavioral spectrum, we uncovered a unique set of individuals in the RCT population who were willing to participate in the RCT but did not proceed to register to the online mental health tool (ie, Nonstarters). These participants were not representative of the community population because they did not encounter the eMH tool at all.

Our study is unique in that we contrasted RCT and general public behavioral patterns for the same eMH program during the same time period, maximizing the comparability of users' eMH experience while minimizing the influence of historical factors. However, some limitations of our analyses warrant consideration. First, our analytic techniques only allowed for a limited number of engagement variables, but many other variables, such as the number of tracking reminders a user sets, could have provided us with a more detailed picture of eMH engagement behavior. Second, we were only able to study those users who actually engaged with the core functionalities of the program repeatedly; therefore, our findings largely reflect individuals motivated to adopt an eMH program. Third, the data reported in this paper were collected between 2011 and 2012 and certainly would have made a timelier contribution then. Technological advances since this time include improvements in interface design and user experience features, such as chatbots, gamification, and virtual reality [15,16]. However, these relatively more high-tech solutions have yet to be fully integrated into the digital mental health landscape [16]. Thus, we believe that the general knowledge and discussion derived from this analysis, which inspected usage behaviors along common metrics such as logins, module usage, and mood monitoring, still bears relevance today. Fourth, the observed effect may be limited in scope. It is possible that the findings presented in this paper only apply to unguided eMH interventions. Usage patterns may differ for eMH programs that provide therapist assistance, which generally facilitates engagement [17]. Last, we did not examine the significance of differences observed across samples. This short paper focused on eMH engagement rather than outcomes, as our goal for this study was to reignite a discussion of the real-world applicability of eMH engagement data derived from RCT findings.

In summary, our findings suggest that eMH engagement in RCTs likely matches the type of eMH engagement in real-world users, but may overestimate the magnitude of such engagement. This could be an important consideration for eMH researchers, designers, and policy makers, as they implement eMH tools after efficacy is established. We recommend that future eMH trials examine whether participant selection and per protocol instructions affect usage behavior, and if so, that due consideration be given to this in implementation planning. Similarly, those planning a wide-scale rollout of new eMH tools should consider which aspects of the original trial may help in promoting usage and whether similar methods can be used in the real world. Ecological validity in eMH engagement science is also relevant to theoretical investigations of how eMH tools improve individuals' well-being. Mechanisms of change established in RCTs must be practicable in the real world for eMH to deliver on its promise of effective and scalable mental health care. Ultimately, the goal of improving eMH engagement science is to set realistic expectations of eMH benefits—both health and economic—and understand how to maximize these. The more accurately we can speak to eMH engagement, the more fruitful both eMH science and policy will be going forward.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures.
[DOCX File , 81 KB - jmir_v23i3e18348_app1.docx]

References

1. Kendall P, Comer J, Chow C. The randomized controlled trial: basics and beyond. In: Comer J, Kendall P, editors. The Oxford Handbook of Research Strategies for Clinical Psychology 9780199793655. New York: Oxford University Press, Incorporated; Apr 23, 2013:40-61.
2. Rothwell PM. Factors that can affect the external validity of randomised controlled trials. PLoS Clin Trials 2006 May;1(1):e9 [FREE Full text] [doi: 10.1371/journal.pctr.0010009] [Medline: 16871331]

3. Baumel A, Kane JM. Examining predictors of real-world user engagement with self-guided eHealth interventions: analysis of mobile apps and websites using a novel dataset. J Med Internet Res 2018 Dec 14;20(12):e11491 [FREE Full text] [doi: 10.2196/11491] [Medline: 30552077]
4. Cavanagh K. Turn on, tune in and (don't) drop out: engagement, adherence, attrition, and alliance with internet-based interventions. In: Bennett-Levy J, Richards D, Farrand P, Christensen H, Griffiths K, Kavanaugh D, et al, editors. Oxford Guide to Low Intensity CBT Interventions (Oxford Guides to Cognitive Behavioural Therapy). New York: Oxford University Press; May 2010:227-233.
5. Sieverink F, Kelders S, Poel M, van Gemert-Pijnen L. Opening the black box of electronic health: collecting, analyzing, and interpreting log data. JMIR Res Protoc 2017 Aug 07;6(8):e156 [FREE Full text] [doi: 10.2196/resprot.6452] [Medline: 28784592]
6. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: A systematic review and meta-analysis. Journal of Affective Disorders 2020 Feb;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
7. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research . . . and three things we can do about them. Psychiatr Serv 2017 May 01;68(5):427-429. [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
8. Ryan C, Bergin M, Wells JS. Theoretical perspectives of adherence to web-based interventions: a scoping review. Int J Behav Med 2018 Feb;25(1):17-29. [doi: 10.1007/s12529-017-9678-8] [Medline: 28730402]
9. Birk MV, Mandryk RL. Improving the efficacy of cognitive training for digital mental health interventions through avatar customization: crowdsourced quasi-experimental study. J Med Internet Res 2019 Jan 08;21(1):e10133 [FREE Full text] [doi: 10.2196/10133] [Medline: 30622095]
10. Holmes NA, van Agteren JE, Dorstyn DS. A systematic review of technology-assisted interventions for co-morbid depression and substance use. J Telemed Telecare 2019 Apr;25(3):131-141. [doi: 10.1177/1357633X17748991] [Medline: 29310532]
11. Blackman KC, Zoellner J, Berrey LM, Alexander R, Fanning J, Hill JL, et al. Assessing the internal and external validity of mobile health physical activity promotion interventions: a systematic literature review using the RE-AIM framework. J Med Internet Res 2013 Oct 04;15(10):e224 [FREE Full text] [doi: 10.2196/jmir.2745] [Medline: 24095951]
12. Christensen H, Griffiths KM, Korten AE, Brittliffe K, Groves C. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website. J Med Internet Res 2004 Dec 22;6(4):e46 [FREE Full text] [doi: 10.2196/jmir.6.4.e46] [Medline: 15631970]
13. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
14. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales (2nd Ed.). Sydney: Psychology Foundation; 1995.
15. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. JMIR Ment Health 2019 Oct 18;6(10):e14166 [FREE Full text] [doi: 10.2196/14166] [Medline: 31628789]
16. Kaonga NN, Morgan J. Common themes and emerging trends for the use of technology to support mental health and psychosocial well-being in limited resource settings: A review of the literature. Psychiatry Res 2019 Nov;281:112594. [doi: 10.1016/j.psychres.2019.112594] [Medline: 31605874]
17. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Review of Neurotherapeutics 2012 Jul;12(7):861-870. [doi: 10.1586/ern.12.63] [Medline: 22853793]

Abbreviations

eMH: e-mental health
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 20.02.20; peer-reviewed by K Mathiasen, B Lo; comments to author 23.06.20; revised version received 30.09.20; accepted 11.12.20; published 11.03.21.

Please cite as:
Sanatkar S, Baldwin P, Huckvale K, Christensen H, Harvey S
e-Mental Health Program Usage Patterns in Randomized Controlled Trials and in the General Public to Inform External Validity Considerations: Sample Groupings Using Cluster Analyses
J Med Internet Res 2021;23(3):e18348
URL: https://www.jmir.org/2021/3/e18348
doi:10.2196/18348
PMID:33704070

Original Paper

Commencement of and Retention in Web-Based Interventions and Response to Prompts and Reminders: Longitudinal Observational Study Based on Two Randomized Controlled Trials

Athanasios Andriopoulos[1*], MSc; Erik M G Olsson[1*], PhD; Ylva Hägg Sylvén[1], MSc; Jonas Sjöström[2], PhD; Birgitta Johansson[3], PhD; Louise von Essen[1], PhD; Helena Grönqvist[1], PhD

[1]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

[2]Department of Informatics and Media, Uppsala University, Visby, Sweden

[3]Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden

[*]these authors contributed equally

Corresponding Author:
Helena Grönqvist, PhD
Department of Women's and Children's Health
Uppsala University
Akademiska sjukhuset
Uppsala, 75185
Sweden
Phone: 46 736236500
Email: helena.gronqvist@kbh.uu.se

Abstract

Background: Web-based interventions are effective for several psychological problems. However, recruitment, adherence, and missing data are challenges when evaluating these interventions.

Objective: This study aimed to describe the use patterns during the commencement phase, possible retention patterns (continuation of data provision), and responses to prompts and reminders among participants in 2 randomized controlled trials (RCTs) evaluating web-based interventions.

Methods: Data on use patterns logged in 2 RCTs aiming to reduce symptoms of anxiety and depression among adult patients recently diagnosed with cancer (AdultCan RCT) and patients with a recent myocardial infarction (Heart RCT) were analyzed. The web-based intervention in the AdultCan trial consisted of unguided self-help and psychoeducation and that in the Heart trial consisted of therapist-supported cognitive behavioral therapy. In total, 2360 participants' use patterns at first log-in, including data collection at baseline (ie, commencement) and at 2 follow-ups, were analyzed. Both the intervention and comparison groups were analyzed.

Results: At commencement, 70.85% (909/1283) and 86.82% (935/1077) of the participants in AdultCan and Heart RCTs, respectively, logged in and completed baseline data collection after receiving a welcome email with log-in credentials. The median duration of the first log-in was 44 minutes and 38 minutes in AdultCan and Heart RCTs, respectively. Slightly less than half of the participants' first log-ins were completed outside standard office hours. More than 80% (92/114 and 103/111) of the participants in both trials explored the intervention within 2 weeks of being randomized to the treatment group, with a median duration of 7 minutes and 47 minutes in AdultCan and Heart RCTs, respectively. There was a significant association between intervention exploration time during the first 2 weeks and retention in the Heart trial but not in the AdultCan trial. However, the control group was most likely to retain and provide complete follow-up data. Across the 3 time points of data collection explored in this study, the proportion of participants responding to all questionnaires within 1 week from the prompt, without a reminder, varied between 35.45% (413/1165) and 66.3% (112/169). After 2 reminders, up to 97.6% (165/169) of the participants responded.

Conclusions: Most participants in both RCTs completed the baseline questionnaires within 1 week of receiving the welcome email. Approximately half of them answered questions at baseline data collection outside office hours, suggesting that the time flexibility inherent in web-based interventions contributes to commencement and use. In contrast to what was expected, the intervention groups generally had lower completion rates than the comparison groups. About half of the participants completed the questionnaires without a reminder, but thereafter, reminders contributed to both baseline and follow-up retention, suggesting

they were effective. Strategies to increase commencement of and retention in eHealth interventions are important for the future development of effective interventions and relevant research.

(J Med Internet Res 2021;23(3):e24590) doi:10.2196/24590

KEYWORDS

log data analysis; use pattern; retention; dropout; attrition; online intervention; online data

Introduction

Background

Web-based interventions are efficient for mental health problems such as symptoms of anxiety, depression, and posttraumatic stress [1-4], with effects lasting up to 3 years after treatment [5-7]. However, studies evaluating web-based interventions struggle with low use, where some participants never log in or commence the intervention at all [8,9] and where retention rates, that is, the continuation of participant data provision, vary between 17% and 98% [10-13]. The problem of low retention has been continuously reported in relation to web-based interventions and research and has even been discussed in terms of *the law of attrition* [13]. Although this problem is not unique to eHealth, the complexity of the field makes attrition almost inevitable, and it is thus important to highlight, measure, and discuss its determinants to be able to improve future eHealth interventions and research, for example, regarding usability, efficacy, and increased acceptability [13]. Disease severity [14], symptoms of anxiety [15], technical issues, lack of motivation, time constraints, the complexity of the intervention, low expectations of its efficacy, compatibility with participants' profiles, and current needs [13,16] have been reported by participants as reasons for noncommencement and low retention [8,17-21]. Demographic variables such as younger age, higher level of education, and female gender are often associated with increased retention in web-based intervention studies [22].

In web-based intervention trials, it is possible to track participants' activities in the intervention by logging use patterns with high precision, including recording every click a participant makes on the web-based platform when working with the intervention and answering questionnaires [23]. Automatized and standardized reminders via emails or text messages are often used to support retention at low cost and with minimal effort [12]. Previous findings indicate that the majority of participants in randomized controlled trials (RCTs) evaluating a web-based intervention perceived reminders as harmless, well accepted, and useful, but the effectiveness of reminders in increasing retention in this type of intervention has seldom been evaluated systematically [24,25]. Log data could be valuable for analyzing the patients' use patterns in web-based interventions and their overall utility [26,27].

Aim and Research Questions

The overall aim of this study is to describe use patterns of participants in 2 RCTs evaluating web-based interventions aimed at reducing symptoms of anxiety and depression in adult patients with cancer (AdultCan trial [28]) and patients who recently had a myocardial infarction (Heart trial [8]) when (1) logging in to the portal for the first time for completing baseline questionnaires (ie, commencement), (2) completing questionnaires at the first and second follow-ups (ie, retention), and (3) responding to prompts and reminders to fill in questionnaires (ie, responses).

Research Questions Regarding Commencement

The research questions regarding commencement were as follows:

- How many potential participants completed the baseline questionnaires and how many left it incomplete?
- Was there a difference in sex or age between those who completed the baseline questionnaires and those who did not?
- How many days after invitation did the participants complete the baseline questionnaires and how many logged in more than once before completing them?
- How long did it take to complete the baseline questionnaires and at what time of the day were the questionnaires completed?

Research Questions Regarding Retention

The research questions regarding retention were as follows:

- How long did the participants explore the intervention after randomization and how many completed follow-up 1 and follow-up 2?
- Was there a difference between those allocated to treatment versus those not regarding completion of follow-up questionnaires?
- Was there an association between exploring activity during the first 2 weeks and completion of follow-up questionnaires?

Research Questions Regarding Prompts and Reminders

The research questions regarding prompts and reminders were as follows:

- How many participants completed questionnaires at follow-up 1 and 2, respectively, after being prompted to do so, and how many responded to questionnaires at follow-ups 1 and 2, respectively, after being reminded one or two times?

Methods

Design

The study had a longitudinal and descriptive correlational design and used secondary data analysis. The primary analysis of the efficacy of the interventions has been reported elsewhere [8,28].

Setting

The Uppsala University Psychosocial Care (U-CARE) program has the overarching goal of promoting psychosocial health among patients struck by somatic diseases and their significant others [23].

The 2 RCTs explored in this study, AdultCan [28] and Heart [8], were conducted via the U-CARE portal (hereafter, portal), a secure web portal developed within U-CARE.

In the AdultCan trial, a stepped-care (consisting of 2 steps) web-based intervention was evaluated. The first step, available for 24 months for each participant, consisted of information, psychoeducation, and self-help material including texts, video lectures, discussion forums, and the possibility for participants to ask questions about cancer and its treatment and get answers from experts. Participants still reported anxiety and depression after access to the first step, and after 1, 4, or 7 months, they were offered a second step consisting of 10 weeks of therapist-supported internet-based cognitive behavioral therapy (iCBT) [28]. Log data collected via the portal during the first step of the intervention were analyzed in this study.

In the Heart trial, a web-based intervention consisting of 14 weeks of therapist-supported iCBT, including self-help material, homework assignments, web-based contact with a therapist, and peer support via a discussion forum, was evaluated. The intervention included 10 modules, for example, behavioral activation, cognitive restructuring, exposure, and problem solving. Participants could choose which modules to work with and receive weekly therapist support [8].

Participants

Log data from 1283 participants in the AdultCan trial and 1077 participants in the Heart trial were analyzed. In the AdultCan trial, the inclusion criteria were patients with newly (within 6 months) diagnosed breast, prostate, or colorectal cancer as well as patients with recurrence of colorectal cancer (within 6 months of diagnosis) at 3 hospitals in Sweden. Exclusion criteria were inability to read and understand Swedish, cognitive disability (eg, dementia or psychosis), a constant need for care (Karnofsky score<40), short expected survival (<3 months), severe depression or suicide risk with regard to answers on the Montgomery-Åsberg Depression Rating Scale-Self-Report (MADRS-S) measure, and participation in a competing clinical trial including prostate cancer patients receiving radiotherapy.

In the Heart trial, inclusion criteria were >7 on one or both of the Hospital Anxiety and Depression Scale (HADS) subscales. Exclusion criteria were scheduled for coronary artery bypass surgery; inability to use a computer, internet, email, or mobile phone; unable to read Swedish; expected to live <1 year; anticipated to show poor compliance (eg, substance abuse); self-reported severe depression or suicidal ideation; MADRS-S item 9>3; and participation in another behavioral intervention trial. Detailed information about the methods used in the AdultCan and Heart RCTs is provided elsewhere [8,28]. In this study, participants who provided informed consent and were added to the portal were considered as participants.

Procedure

In both studies, participants who self-reported symptoms of anxiety and/or depression above the cut-off >7 on any of the subscales of the HADS were randomized to either the treatment group or the control group. In the AdultCan trial, those scoring below the cut-off on both subscales were assigned to a reference group that was followed longitudinally. Details of the procedure at commencement and data collection in the AdultCan and Heart trials are presented in Table 1.

Table 1. The procedure at commencement and follow-up data collection in AdultCan and Heart trials.

Phase, Studies	AdultCan trial	Heart trial
Commencement	• Eligible persons were informed about the study at a regular hospital visit or by telephone within 6 months after being diagnosed with cancer. • After providing written informed consent, participants received a welcome email with log-in credentials to the portal for baseline questionnaires. • Participants were informed that if all baseline questions were not answered within 24 h, they would have to restart from scratch. • If participants did not complete the 14 baseline questionnaires within 7 days, they received a reminder via SMS and email. • If participants did not complete the baseline questionnaires within 14 days, they received a second reminder via SMS and email. • If participants had still not completed the baseline questionnaires 30 days after the prompt, study personnel contacted them, if possible, by telephone and reminded them to respond to the questionnaires. • Participants scoring above the cut-off on HADS[a] were randomized to the treatment or control group in the portal. The log-in session where a participant is randomized is called the "randomization session." • Participants scoring below the cut-off were assigned to the reference group and were asked to answer questionnaires at selected time points. • Participants randomized to the treatment group got immediate access to the first step of the intervention via the portal.	• Eligible persons were informed about the study at a regular hospital visit shortly after discharge from the hospital after an MI[b]. Potential participants were then contacted again 8 weeks after the MI by the study staff via telephone. • After providing written informed consent, participants received a welcome email with log-in credentials to the portal for baseline questionnaires. • Participants were informed that if all baseline questions were not answered within 24 h, they would have to restart from scratch. • If participants did not complete the 13 baseline questionnaires within 7 days, they received a reminder via SMS and email. • If participants did not complete the baseline questionnaires within 14 days, they were reminded by study personnel via telephone. • Participants scoring above the cut-off on HADS were randomized to the treatment or control group in the portal. The log-in session where a participant is randomized is called the "randomization session." • Participants randomized to the treatment group got immediate access to the intervention via the portal.
Retention, prompts, and reminders	• Follow-up 1: 2 weeks after randomization, participants were asked to complete 1 (control group) or 2 (treatment group) questionnaires. • Follow-up 2: 1 month after randomization, participants were asked to complete 4 (reference group), 8 (control group), or 10 (treatment group) questionnaires. • At follow-up 1 and 2 participants were prompted via email and SMS to log in and complete the questionnaires. • If participants did not complete the questionnaires within 7 days after the prompt, they received a first reminder via SMS and email. • If participants did not complete the questionnaires within 12 days after the prompt, they received a second reminder via SMS and email.	• Follow-up 1: 5 weeks after randomization, participants were asked to complete 4 questionnaires. • Follow-up 2: 14 weeks after randomization, participants were asked to complete 14 questionnaires. • At follow-up 1 and 2 participants were prompted via email and SMS to log in and complete questionnaires. • If participants did not complete the questionnaires within 7 days after the prompt, they received a first reminder via SMS and email. • If participants did not complete the questionnaires within 14 days after the prompt, study personnel contacted them via telephone and reminded them to respond to the questionnaires.

[a]HADS: Hospital Anxiety and Depression Scale.

[b]MI: myocardial infarction.

Participants logged in to the portal with double authentication, entering username, personal password, and a temporary 5-digit code that they received in an SMS. A log-in session ended when a participant logged out of the portal or was inactive for more than 20 minutes.

In the Heart trial, at the second reminder, participants were offered to use paper forms to answer questionnaires at follow-up 1 and 2, which 27 and 46 participants did at the 2 follow-ups, respectively. Thus, the web-based completion rate is only a part of the total completion rate in the Heart trial.

Data and Data Collection

Log data from the full duration of the AdultCan and Heart trials were collected from April 16, 2013, to April 28, 2017, and were exported from the portal by a system developer. The data were reviewed by a second system developer. In addition, the researchers performed random checks and reviewed any inconsistencies.

Log data refers to records of real-time actions performed by each user, and mouse clicks and keyboard strokes are logged as user actions with time stamps. In this study, log data at commencement, during the 2-week period following commencement, and at 2 consecutive follow-up time points within the RCTs AdultCan [28] and Heart [8] were collected via the secure portal developed within U-CARE.

Variables

The variables used to answer the research questions are presented in Table 2.

Table 2. Variables used in the study.

Phase and study variables measured	Value
Commencement	
Participant commencing answering questionnaires at baseline	y/n[a]
Participant completing answering questionnaires at baseline	y/n
Time from the welcome email sent from the portal with log-in credentials to participant's first log-in	d[b];h:min
Duration of participant first log-in	min:s
Time of the day when the participant first logged in	h:min
Day of week when the participant first logged in	Mo-Su[c]
Whether the participant's first log-in ended with a click by the participant on the log-out button or if the participant was automatically logged out after being inactive (passive log-out)	y/n
Retention	
Participant explored the intervention in randomization session (treatment group only)	y/n
Participant explored the intervention within 14 days after randomization (treatment group only)	y/n
Length of time the participant explored the intervention within 14 days after randomization (treatment group only)	min:s
Participant completed all questionnaires at follow-up 1 and 2	y/n
Response to prompts and reminders	
Number of prompts and/or reminders sent to participants at the 3 data collection time points	0-2

[a]y/n: yes or no.

[b]d: day.

[c]Mo-Su: Monday to Sunday.

The following portal activities were defined as exploring the intervention: any click in the library, forum, chat, diary, FAQ, ask an expert, using the internal message system, and the iCBT program.

Self-reported demographical data were collected at baseline.

Missing Data

Missing data were mostly because messages, such as prompts and reminders, from the portal were not logged properly, as a result of a temporary technical error in the early phase after launching the studies. The welcome emails with log-in credentials were erroneously logged for 5 and 7 participants in the respective studies, and reminders to log in to the portal to answer questionnaires at baseline were erroneously or insufficiently logged for 118 and 50 participants from the AdultCan and Heart RCTs, respectively, with missing data as a result. The corresponding figures for the first follow-up were 76 and 19, and for the second follow-up, 112 and 18. In the Heart trial, 68, 24, and 42 participants were not reached by telephone for reminders at baseline, first, and second follow-up, respectively. The country of birth was not reported by one participant. When investigating exploration, 10 participants in the AdultCan trial and 7 in the Heart trial had missing data.

Statistical Analysis

Descriptive statistics were used to examine and report all variables. Medians were used when the frequency distributions were skewed. Pearson chi-square test was used to examine potential differences between the numbers of participants exploring the intervention among participants who completed the baseline (completers) and those who did not complete the baseline (noncompleters) in the respective study groups. The Mann-Whitney U-test was used to examine potential associations between time used to explore the intervention and if participants completed the data collections in the respective studies. Actual *P* values are reported. All analyses were based on complete data, that is, no imputations were performed.

Data were analyzed using IBM SPSS Statistics V25.0 and STATA v 15.1.

Results

Patient Characteristics

Participants in the AdultCan and Heart trials who completed baseline questionnaires had a mean age of 61 years (SD 10.6) and 62 years (SD 8.1), respectively, and at least 90.29% (1665/1844) were born in Sweden, and more than 44.74% (825/1844) had some university education. In the AdultCan trial, the proportion of female participants was more than double that in the Heart trial (for more details, Table 3).

Table 3. Characteristics and commencement data of participants.

Characteristics and commencement data	AdultCan trial		Heart trial	
	Completed BL questionnaire[a] (n=909)	Did not complete BL questionnaire(n=374)	Completed BL questionnaire(n=935)	Did not complete BL questionnaire(n=142)
Age (years), mean (SD)	61.3 (11)	62.5 (11)	62.2 (8)	62.4 (9)
Women, n (%)	525 (57.8)	207 (55.3)	220 (23.5)	34 (23.9)
Born in Sweden, n (%)	819 (90.1)	N/A[b]	846 (90.5)	N/A
Highest level of education, n (%)				
Elementary school	184 (20.2)	N/A	190 (20.3)	N/A
High school	296 (32.6)	N/A	349 (37.3)	N/A
University ≤3 years	193 (21.2)	N/A	185 (29.8)	N/A
University >3 years	236 (26.0)	N/A	211 (22.6)	N/A
Study group, n (%)				
Reference group	664 (72.9)	N/A	696 (74.4)	N/A
Control group	121 (13.3)	N/A	122 (13.0)	N/A
Treatment group	124 (13.6)	N/A	117 (12.5)	N/A
Log-ins for noncompleters BL questionnaire, n (%)				
≥1 log-ins	N/A	83 (22.2)	N/A	33 (23.2)
No log-ins	N/A	291 (77.8)	N/A	109 (76.8)
Time to first log-in (d;h:min), median (range)	6;11:59 (0;00:01-59;04:44)	N/A	3;21:57 (0;00:02-18;10:32)	N/A
First log-in duration (min:s), median (range)	44:08 (00:31-180:06)	N/A	37:43 (00:19-326:48)	N/A

[a]BL: baseline.

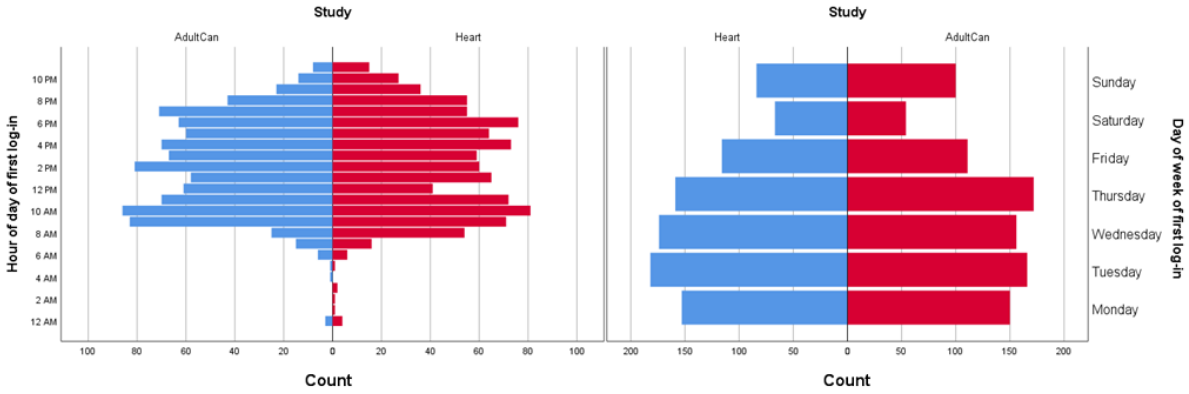
[b]N/A: not applicable

Commencement

A total of 70.85% (909/1283) of participants in the AdultCan trial and 86.82% (935/1077) of participants in the Heart trial completed all questionnaires at baseline (Table 3). Moreover, 6.47% (83/1283) of the participants in the AdultCan trial did not complete all baseline questionnaires after being logged in at least once, and 22.68% (291/1283) of participants did not log in at all. In the Heart trial, these numbers were approximately half (Table 3). The median response time, from receiving the welcome email to the first log-in, was slightly more than 6 days in the AdultCan trial and almost 4 days in the Heart trial. Of those who completed all questionnaires at baseline, 24.9% (226/909) in the AdultCan trial and 23.7% (222/935) in the Heart trial logged in more than once before completing the baseline questionnaires. Of those who completed all questionnaires at baseline, 73.0% (664/909) in the AdultCan trial and 74.4% (696/935) in the Heart trial were allocated to the reference group and not randomized. The median duration of the first log-in, from the first to the last click, was 44 minutes and 38 minutes for the AdultCan and Heart trials, respectively. At baseline, 54.3% (494/909) of the participants in the AdultCan trial and 35.8% (335/935)of the participants in the Heart trial logged out of the portal by a click. No differences in age or gender were found between participants who completed the baseline questionnaires and those who did not.

The times when the participants logged in for the first time are illustrated in Figure 1. In the AdultCan trial, 54.8% (498/909) of the first-time log-ins were on weekdays between 8 AM and 5 PM, representing normal office hours. The corresponding figure for the Heart trial was 52.2% (488/935).

Figure 1. The time of day and day of week for participants' first log-in.



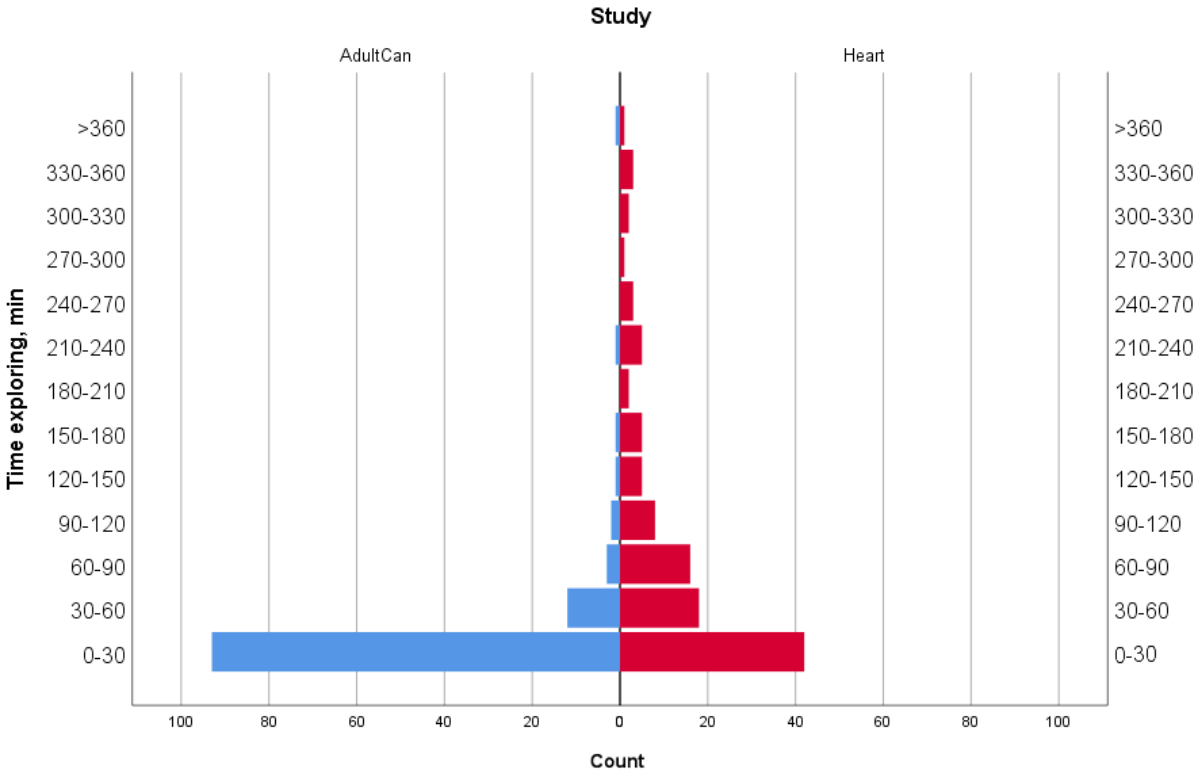
Retention

A total of 73.7% (84/114) of those randomized to treatment in the AdultCan trial and 70.3% (78/111) of those in the Heart trial explored the intervention within the session when they completed baseline questionnaires and were randomized to the treatment group. Thereafter, within a 14 day-period after randomization, separate from the randomization session, 29.8% (34/114) of the participants in the AdultCan trial and 72.1% (80/111) of the participants in the Heart trial explored the intervention at least once. The median total time participants were exploring the intervention during the first 14 days after randomization was 7 minutes for the AdultCan trial and 47 minutes for the Heart trial (Table 4). Figure 2 provides a detailed description of the distribution of total time spent exploring the intervention.

Table 4. Number of participants in the treatment group exploring the intervention within 14 days after randomization and their time spent on exploring.

Measures of exploration	AdultCan trial (n=144)	Heart trial (n=111)
Exploring the intervention, n (%)	92 (80.7)	103 (92.8)
Total time spent exploring the intervention (min:s), median (range)	6:58 (00:01-461:40)	47:14 (00:22-486:06)

Figure 2. Total time spent in exploring the intervention within 14 days after randomization by number of participants in the treatment group in AdultCan and Heart trials.



Overall, the questionnaires were completed using the web-based platform by 44%-85% of the participants in the different study groups (treatment, control, and reference group) at the 2 follow-ups (Table 5). In the Heart trial, more participants in the control group compared with the treatment group completed the questionnaires using the web-based platform both at follow-up 1 and follow-up 2. No differences between the groups were found in the AdultCan trial at follow-up 1, but at follow-up 2, the reference group completed the questionnaires the most and the treatment group the least (Table 5).

Table 5. Web-based completion rates for the follow-ups for the AdultCan and Heart trials.

Follow-up completion	AdultCan trial				Heart trial		
	Reference, n (%)	Control (n=121), n (%)	Treatment (n=124), n (%)	*P* value[a]	Control (n=121), n (%)	Treatment (n=117), n (%)	*P* value[a]
Completing FU[b]1	N/A[c]	82 (67.7)	82 (66.1)	.79	104 (85.2)	80 (68.3)	.002
Completing FU2	558 (84)	94 (77.7)	90 (72.6)	.005	96 (78.7)	52 (44.4)	<.001

[a]*P* values from chi-square tests.

[b]FU: follow-up.

[c]N/A: not applicable.

When combining total exploration time, the first 14 days with retention, completing follow-up 1 and follow-up 2 were positively associated with the exploration time for the first 14 days after randomization in the Heart trial but not with any of the follow-ups in the AdultCan trial (Table 6). When dividing participants in the Heart trial into an active and a passive treatment group, based on a median split in exploration time (median 47 minutes and 14 seconds), 55% (30/55) in the passive treatment group compared with 80% (45/56) in the active group and 85.2% (104/122) in the control group completed follow-up 1. At follow-up 2, the corresponding figures were 31% (17/55), 55% (31/56), and 62.3% (96/122).

Table 6. Total time exploring the intervention during the first 14 days cross-tabled with completion of follow-up measures.

Follow-up	AdultCan trial (n=114)				Heart trial (n=111)			
	Completed web-based follow-up	Did not complete web-based follow-up	Diff.[a]	*P* value[b]	Completed using web-based intervention	Did not complete web-based follow-up	Diff.	*P* value[b]
Follow-up 1								
n (%)	75 (66)	39 (34)			75 (68)	36 (32)		
Total time spent exploring the intervention (min:s), median (range)	6:35 (00:01-461:40)	07:20 (00:01-76:05)	−0:45	.55	70:37 (00:23-353:40)	23:53 (00:22-486:05)	46:44	<.001
Follow-up 2								
n (%)	83 (73)	31 (27)			48 (43)	63 (57)		
Total time spent exploring the intervention (min:s), median (range)	7:34 (00:01-461:40)	4:08 (00:02-75:05)	3:26	.46	76:16 (00:22-353:40)	34:03 (00:23-486:05)	42:13	.005

[a]Diff.: difference in median times between those who completed web-based follow-ups and those who did not.

[b]*P* values from Mann-Whitney U-tests.

Response to Prompts and Reminders

Across the 3 data collection time points explored in this study, the proportion of participants responding within 7 days from the prompt without a reminder was between 36% and 66%. Within 5 days of the first reminder, sent out via SMS and email, an additional 40%-86% of the remaining participants responded. In the AdultCan trial, the second reminder, sent via SMS and email, generated between 36% and 50% additional responses from the remaining participants. In the Heart trial, the second reminder, via telephone, generated 34%-69% additional responses among those who had not responded so far (Table 7).

Table 7. Responses to reminders at baseline and the 2 follow-ups.

Time of response	AdultCan						Heart					
	Baseline		FU[a]1		FU2		Baseline		FU1		FU2	
	N	n (%)	N	n (%)	N	n (%)	N	n (%)	N	n (%)	N	n (%)
Response after prompt, n (% of all)	1165	413 (35.5)	169	112 (66.2)	797	442 (55.5)	959	550 (57.3)	196	109 (64.5)	179	74 (41.3)
Response to the first reminder, n (% of remaining)	752	299 (39.8)	57	49 (86.0)	355	151 (42.5)	409	209 (51.1)	87	50 (57.5)	105	44 (41.9)
Response to the second reminder, n (% of remaining)	453	162 (35.8)	8	4 (50.0)	204	99 (48.5)	200	137 (68.5)	37	16 (43.2)	61	23 (37.7)
No response, n (% of all)	1165	291 (25.0)	169	4 (2.4)	797	105 (13.2)	959	63 (6.6)	196	21 (10.7)	179	38 (21.2)

[a]FU: follow-up.

Discussion

Principal Findings

The results show that at commencement, most recruited and consenting participants logged in and completed the baseline questionnaires. Most nonresponders did not log in at all. In contrast to previous studies [29] that have indicated that age and gender are related to attrition, no difference in gender or age was found between the participants who finished baseline and those who did not. This may be partly because of age heterogeneity. Fewer participants in the AdultCan trial than those in the Heart trial completed the baseline questionnaires. This could have many reasons, such as recruitment procedure, intervention type, disease severity, and so on. Most participants in the AdultCan trial were undergoing active cancer treatment at the time of inclusion, whereas the focus of participants in the Heart trial was on secondary prevention. Most participants who completed the baseline questionnaires completed the questionnaire within 1 week of receiving the welcome email with log-in credentials from the portal.

One argument for using web-based interventions is that they can be accessed at any time. Although most participants had their first log-in on weekdays and during the day, 45%-48% of the participants chose to commence the studies outside common office hours when face-to-face psychological support is usually not offered. The log-in times were similar in the 2 studies regarding time of day and day of the week and also similar to what has been reported in other studies [30]. It is known that the time of day and the day of week people prefer to answer surveys are related to sociodemographic and health characteristics [31]. As internet interventions are flexible in time, they may be able to reach patients in need at convenient times.

Most participants opened at least one item of the intervention directly after being randomized to treatment. Furthermore, in the Heart trial, 72% of the participants explored the intervention in separate sessions during the following 14 days. This was more than that in the AdultCan trial. Median time logged in during the first 14 days was also longer in the Heart trial than in the AdultCan trial. This was expected owing to the intervention formats, as the Heart trial was a therapist-supported iCBT intervention, whereas the AdultCan trial offered self-help psychoeducation without individual support during the first 2 weeks examined in this study. In addition, therapist-supported iCBT was restricted to 10 weeks, whereas self-help psychoeducation in the AdultCan trial was available for 24 months. However, the overall intervention use over the first 2 weeks was relatively low. Persuasive features such as feedback have been suggested to increase use [32] and were available in the Heart intervention. However, the participants had to log in without any specific prompts to notice the feedback.

Most participants (66%-85%) were retained in the studies and answered the follow-up questionnaires. When comparing completion rates between the study groups, the control group in the Heart trial had a higher rate than that in the treatment group at both follow-ups. A similar pattern was evident in the AdultCan trial at follow-up 2, where the reference group had the highest completion rate and the treatment group the lowest. Although the more active treatment participants in the Heart study also had a higher completion rate than those who were less active, the active participants were still less likely to complete the follow-ups than the control group. This was an unexpected finding. It may be that participants felt obliged to contribute to a certain amount and that those participating in the intervention thought they had filled their quota even before the follow-up questionnaires. To the best of our knowledge, there are no previous systematically summarized studies reflecting on such patterns.

Prompts and reminders for completing questionnaires were sent via SMS and email. Most participants answered the questionnaires after the prompt without any reminders. However, the following 2 consecutive reminders were useful in increasing the response rates, not only when executed via telephone calls but also via SMS and email. The results are in line with previous research showing that reminders contribute to the overall response rate [33] and that participants find reminders acceptable and useful [25]. In the Heart trial, participants were offered paper forms as a secondary response alternative at the second reminder, which should be considered when interpreting the sometimes very low retention rates.

Strengths and Limitations

The log data collected for this study allowed for a unique possibility of exploring these aspects that are important for the success of web-based interventions. Using participants in 2 web-based intervention studies gave us a large sample size of 2360 participants. There are several differences between the studies, making them difficult to compare; hence, they are described as separate cases with few comparisons. However, the results were similar, and the 2 cases provided cumulative information for the exploration of use patterns. Another strength is that both studies recruited clinically and consecutively, resulting in a sample from all patients, not only self-selected highly motivated participants in web-based interventions. We believe that a more detailed log data on participants' use patterns could improve the development of future web-based interventions.

The second reminder at follow-up 1 and 2 in the Heart trial was made by telephone. To maximize responses, participants were offered to answer the questionnaire by pen and paper if they were reluctant to log in and answer via the portal. However, as this study focuses on use patterns, the questionnaires filled in by pen and paper answers were not considered. However, they have been reported in the Heart main study outcomes [8].

All data were logged using the portal. Researchers decided what to log beforehand but did not influence the data during data collection. There are some missing data, especially regarding reminders, and the data were not logged properly when the study commenced. However, the quality of the data extracted and analyzed in this study was high and reliable.

Conclusions

Although use patterns differed slightly between the 2 studies, some general conclusions can be drawn. Most people who consented to participate in the study commenced by completing the baseline questionnaires within 1 week. Although many participants answered the questionnaires on the portal during office hours, approximately half of them did so during the weekend or in the evenings, suggesting that flexibility contributes to commencement and use. Participants in the study treatment groups tended to have lower completion rates for the follow-up questionnaires than those in the control or reference groups. This unexpected finding would be interesting for further investigations. Reminders were important to improve the completion rate of questionnaires at baseline and at follow-up. A second reminder was effective in increasing the completion rate. To summarize, our results show that log data provide a rich source of information for a better understanding of use patterns in web-based intervention and retention in eHealth trials. We found that commencement and retention are related to, among other things, flexibility, study design features, and reminders. Our results not only largely support previous findings but also indicate some unexpected user patterns to be investigated further. Refined logging and complementary interviews could potentially provide an even better understanding of these behavioral patterns. As we learn more about users' detailed behaviors, we need improved intervention design and data collection that use the strengths and weaknesses of the internet format.

Acknowledgments

This study was funded by the Swedish Research Council for the strategic research program U-CARE. The funder was not involved in the review or approval of the manuscript.

Conflicts of Interest

None declared.

References

1. Cernvall M, Carlbring P, Ljungman L, Ljungman G, von Essen L. Internet-based guided self-help for parents of children on cancer treatment: a randomized controlled trial. Psychooncology 2015 Sep;24(9):1152-1158 [FREE Full text] [doi: 10.1002/pon.3788] [Medline: 25728688]
2. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
3. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
4. Kuester A, Niemeyer H, Knaevelsrud C. Internet-based interventions for posttraumatic stress: A meta-analysis of randomized controlled trials. Clin Psychol Rev 2016 Feb;43:1-16. [doi: 10.1016/j.cpr.2015.11.004] [Medline: 26655959]
5. Andersson G, Hesser H, Veilord A, Svedling L, Andersson F, Sleman O, et al. Randomised controlled non-inferiority trial with 3-year follow-up of internet-delivered versus face-to-face group cognitive behavioural therapy for depression. J Affect Disord 2013 Dec;151(3):986-994. [doi: 10.1016/j.jad.2013.08.022] [Medline: 24035673]
6. Ruwaard J, Schrieken B, Schrijver M, Broeksteeg J, Dekker J, Vermeulen H, et al. Standardized web-based cognitive behavioural therapy of mild to moderate depression: a randomized controlled trial with a long-term follow-up. Cogn Behav Ther 2009 Feb;38(4):206-221. [doi: 10.1080/16506070802408086] [Medline: 19221919]

7. Cernvall M, Carlbring P, Wikman A, Ljungman L, Ljungman G, von Essen L. Twelve-Month Follow-Up of a Randomized Controlled Trial of Internet-Based Guided Self-Help for Parents of Children on Cancer Treatment. J Med Internet Res 2017 Jul 27;19(7):e273 [FREE Full text] [doi: 10.2196/jmir.6852] [Medline: 28751300]
8. Norlund F, Wallin E, Olsson EMG, Wallert J, Burell G, von Essen L, et al. Internet-Based Cognitive Behavioral Therapy for Symptoms of Depression and Anxiety Among Patients With a Recent Myocardial Infarction: The U-CARE Heart Randomized Controlled Trial. J Med Internet Res 2018 Mar 08;20(3):e88 [FREE Full text] [doi: 10.2196/jmir.9710] [Medline: 29519777]
9. Wangberg SC, Bergmo TS, Johnsen JK. Adherence in Internet-based interventions. Patient Prefer Adherence 2008 Feb 02;2:57-65 [FREE Full text] [Medline: 19920945]
10. Melville KM, Casey LM, Kavanagh DJ. Dropout from Internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
11. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of Interventions Delivered Online and Through Mobile Phones for People Who Experience Severe Mental Health Problems: A Systematic Review. J Med Internet Res 2016 May 31;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]
12. Gandy M, Fogliati VJ, Terides MD, Johnston L, Nicholson Perry K, Newall C, et al. Short message service prompts for skills practice in Internet-delivered cognitive behaviour therapy for chronic pain - are they feasible and effective? Eur J Pain 2016 Sep;20(8):1288-1298. [doi: 10.1002/ejp.853] [Medline: 27072189]
13. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
14. Beatty L, Kemp E, Binnion C, Turner J, Milne D, Butow P, et al. Uptake and adherence to an online intervention for cancer-related distress: older age is not a barrier to adherence but may be a barrier to uptake. Support Care Cancer 2017 Dec;25(6):1905-1914. [doi: 10.1007/s00520-017-3591-1] [Medline: 28155018]
15. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol Med 2015 Oct;45(13):2717-2726. [doi: 10.1017/S0033291715000665] [Medline: 25881626]
16. Alfonsson S, Olsson E, Hursti T. Motivation and Treatment Credibility Predicts Dropout, Treatment Adherence, and Clinical Outcomes in an Internet-Based Cognitive Behavioral Relaxation Program: A Randomized Controlled Trial. J Med Internet Res 2016 Mar 08;18(3):e52 [FREE Full text] [doi: 10.2196/jmir.5352] [Medline: 26957354]
17. van Ballegooijen W, Cuijpers P, van Straten A, Karyotaki E, Andersson G, Smit JH, et al. Adherence to Internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. PLoS One 2014;9(7):e100674 [FREE Full text] [doi: 10.1371/journal.pone.0100674] [Medline: 25029507]
18. Wallin EEK, Mattsson S, Olsson EMG. The Preference for Internet-Based Psychological Interventions by Individuals Without Past or Current Use of Mental Health Treatment Delivered Online: A Survey Study With Mixed-Methods Analysis. JMIR Ment Health 2016 Jun 14;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]
19. Whitfield G, Williams C. If the Evidence Is So Good – Why Doesn't Anyone Use Them? A National Survey af the Use of Computerized Cognitive Behaviour Therapy. Behav Cogn Psychother 2004;32(1):57-65. [doi: 10.1017/S1352465804001031]
20. Kaltenthaler E, Brazier J, De NE, Tumur I, Ferriter M, Beverley C, et al. Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation. Health Technol Assess 2006 Sep;10(33):1-168 [FREE Full text] [doi: 10.3310/hta10330] [Medline: 16959169]
21. Alfonsson S, Maathz P, Hursti T. Interformat reliability of digital psychiatric self-report questionnaires: a systematic review. J Med Internet Res 2014 Dec 03;16(12):e268 [FREE Full text] [doi: 10.2196/jmir.3395] [Medline: 25472463]
22. El Alaoui S, Ljótsson B, Hedman E, Kaldo V, Andersson E, Rück C, et al. Predictors of Symptomatic Change and Adherence in Internet-Based Cognitive Behaviour Therapy for Social Anxiety Disorder in Routine Psychiatric Care. PLoS One 2015;10(4):e0124258 [FREE Full text] [doi: 10.1371/journal.pone.0124258] [Medline: 25893687]
23. Sjöström J, von Essen L, Grönqvist H. The Origin and Impact of Ideals in eHealth Research: Experiences From the U-CARE Research Environment. JMIR Res Protoc 2014 May 23;3(2):e28 [FREE Full text] [doi: 10.2196/resprot.3202] [Medline: 24860071]
24. Kannisto KA, Adams CE, Koivunen M, Katajisto J, Välimäki M. Feedback on SMS reminders to encourage adherence among patients taking antipsychotic medication: a cross-sectional survey nested within a randomised trial. BMJ Open 2015 Nov 09;5(11):e008574 [FREE Full text] [doi: 10.1136/bmjopen-2015-008574] [Medline: 26553830]
25. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009 May 14;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
26. Sieverink F, Kelders S, Poel M, van Gemert-Pijnen L. Opening the Black Box of Electronic Health: Collecting, Analyzing, and Interpreting Log Data. JMIR Res Protoc 2017 Aug 07;6(8):e156 [FREE Full text] [doi: 10.2196/resprot.6452] [Medline: 28784592]
27. Sjöström J, Rahman M, Rafiq A, Lochan R, Ågerfalk P. Respondent Behavior Logging: An Opportunity for Online Survey Design. In: vom Brocke J, Hekkala H, Ram S, Rossi M, editors. Design Science at the Intersection of Physical and Virtual Design. DESRIST 2013. Lecture Notes in Computer Science, vol 7939. In Berlin Heidelberg: Springer; 2013:511-518.

28. Hauffman A, Alfonsson S, Bill-Axelson A, Bergkvist L, Forslund M, Mattsson S, et al. Cocreated internet-based stepped care for individuals with cancer and concurrent symptoms of anxiety and depression: Results from the U-CARE AdultCan randomized controlled trial. Psychooncology 2020 Dec;29(12):2012-2018 [FREE Full text] [doi: 10.1002/pon.5489] [Medline: 32691455]
29. Kannisto KA, Korhonen J, Adams CE, Koivunen MH, Vahlberg T, Välimäki MA. Factors Associated With Dropout During Recruitment and Follow-Up Periods of a mHealth-Based Randomized Controlled Trial for Mobile.Net to Encourage Treatment Adherence for People With Serious Mental Health Problems. J Med Internet Res 2017 Feb 21;19(2):e46 [FREE Full text] [doi: 10.2196/jmir.6417] [Medline: 28223262]
30. Harjumaa M, Absetz P, Ermes M, Mattila E, Männikkö R, Tilles-Tirkkonen T, et al. Internet-Based Lifestyle Intervention to Prevent Type 2 Diabetes Through Healthy Habits: Design and 6-Month Usage Results of Randomized Controlled Trial. JMIR Diabetes 2020 Aug 11;5(3):e15219 [FREE Full text] [doi: 10.2196/15219] [Medline: 32779571]
31. Mindell J, Aresu M, Becares L, Tolonen H. Representativeness of participants in a cross-sectional health survey by time of day and day of week of data collection. Eur J Public Health 2011 Sep 29;22(3):364-369. [doi: 10.1093/eurpub/ckr093]
32. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
33. Van Mol C. Improving web survey efficiency: the impact of an extra reminder and reminder content on web survey response. Int J Soc Res Method 2016 May 17;20(4):317-327. [doi: 10.1080/13645579.2016.1185255]

Abbreviations

HADS: Hospital Anxiety and Depression Scale
iCBT: internet-based cognitive behavioral therapy
MADRS-S: Montgomery-Åsberg Depression Rating Scale-Self-Report
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 25.09.20; peer-reviewed by SFM Pizzoli, S Lambert; comments to author 07.11.20; revised version received 11.12.20; accepted 16.01.21; published 12.03.21.

Please cite as:
Andriopoulos A, Olsson EMG, Hägg Sylvén Y, Sjöström J, Johansson B, von Essen L, Grönqvist H
Commencement of and Retention in Web-Based Interventions and Response to Prompts and Reminders: Longitudinal Observational Study Based on Two Randomized Controlled Trials
J Med Internet Res 2021;23(3):e24590
URL: https://www.jmir.org/2021/3/e24590
doi:10.2196/24590
PMID:33709937

Original Paper

Immersive Virtual Reality and Ocular Tracking for Brain Mapping During Awake Surgery: Prospective Evaluation Study

Morgane Casanova[1], MSc; Anne Clavreul[2,3], PhD; Gwénaëlle Soulard[2,3], MSc; Matthieu Delion[2], MD, PhD; Ghislaine Aubin[2], PhD; Aram Ter Minassian[4], MD, PhD; Renaud Seguier[1], PhD; Philippe Menei[2,3], MD, PhD

[1]Équipe Facial Analysis Synthesis & Tracking Institue of Electronics and Digital Technologies, CentraleSupélec, Rennes, France

[2]Département de Neurochirurgie, Centre hospitalier universitaire d'Angers, Angers, France

[3]Centre de Recherche en Cancérologie et Immunologie Nantes Angers, Université d'Angers, Centre hospitalier universitaire d'Angers, Angers, France

[4]Département d'Anesthésie-Réanimation, Centre hospitalier universitaire d'Angers, Angers, France

Corresponding Author:
Philippe Menei, MD, PhD
Département de Neurochirurgie
CHU d'Angers
4 rue Larrey
Angers
France
Phone: 33 241 354822
Email: phmenei@chu-angers.fr

Abstract

Background: Language mapping during awake brain surgery is currently a standard procedure. However, mapping is rarely performed for other cognitive functions that are important for social interaction, such as visuospatial cognition and nonverbal language, including facial expressions and eye gaze. The main reason for this omission is the lack of tasks that are fully compatible with the restrictive environment of an operating room and awake brain surgery procedures.

Objective: This study aims to evaluate the feasibility and safety of a virtual reality headset equipped with an eye-tracking device that is able to promote an immersive visuospatial and social virtual reality (VR) experience for patients undergoing awake craniotomy.

Methods: We recruited 15 patients with brain tumors near language and/or motor areas. Language mapping was performed with a naming task, DO 80, presented on a computer tablet and then in 2D and 3D via the VRH. Patients were also immersed in a visuospatial and social VR experience.

Results: None of the patients experienced VR sickness, whereas 2 patients had an intraoperative focal seizure without consequence; there was no reason to attribute these seizures to virtual reality headset use. The patients were able to perform the VR tasks. Eye tracking was functional, enabling the medical team to analyze the patients' attention and exploration of the visual field of the virtual reality headset directly.

Conclusions: We found that it is possible and safe to immerse the patient in an interactive virtual environment during awake brain surgery, paving the way for new VR-based brain mapping procedures.

Trial Registration: ClinicalTrials.gov NCT03010943; https://clinicaltrials.gov/ct2/show/NCT03010943.

(J Med Internet Res 2021;23(3):e24373) doi:10.2196/24373

KEYWORDS

virtual reality; eye tracking; brain mapping; awake surgery; visuospatial cognition; nonverbal language; mobile phone

Introduction

Brain mapping by direct electrical stimulation (DES) during awake craniotomy is currently a standard procedure that reduces the risk of permanent neurological deficits and increases the extent of tumor resection and the success of epilepsy surgery [1]. This technique aims to temporarily inactivate a discrete brain area using DES while the patient performs a task. If the patient's performance in the task decreases during inactivation, then the region of the brain explored is considered eloquent for the task and is preserved.

Verbal language, which is controlled by the dominant hemisphere, is widely mapped in this way [2]. Other cognitive functions, such as visuospatial cognition and nonverbal language, including facial expressions and eye gaze, which play an important role in social interaction, have been explored by only a few groups [3]. One of the main reasons for the lack of mapping these functions is the difficulty in adapting classic bedside neuropsychological tasks to awake surgery conditions. In particular, the patient must give an unambiguous answer within 5 seconds, which is the maximum duration of DES. Therefore, there is a need for new tools, allowing complex neuropsychological evaluations, that are compatible with the restrictive environment of an operating room and awake brain surgery procedures.

A few years ago, we began exploring the feasibility of testing cognitive functions during awake craniotomy by immersing the patient in virtual situations with a virtual reality headset (VRH). We have developed several different approaches using different types of headsets and software. The first virtual reality (VR) tasks were developed with the aim of preventing postoperative hemianopsia and unilateral neglect [4]. However, the risk of VR sickness or VRH-induced seizures raised concerns for all screen-based video games [5-8]. Therefore, we performed an initial study evaluating the tolerance of a wireless, low-cost, high-quality, and customizable device: the Samsung Gear VR combined with a Samsung S7 smartphone [9]. This trial, on 30 patients, showed that VRH use and immersive virtual experiences were both feasible and safe for patients undergoing awake craniotomy and brain mapping using DES. Various VR experiences were tested, including a picture-naming task, DO 80 [10] and, a social VR task, vTime, simulating virtual social interactions with an avatar piloted by a neuropsychologist, who also wore a VRH [11]. Toward the end of this safety study, a new VRH with higher performance, the HTC VIVE (HTC Corporation), including an eye-tracking device, was released. Therefore, we decided to prolong the study, including 15 more patients using this new device, to analyze the feasibility of using eye tracking during awake craniotomy. Furthermore, during the initial trial, we experienced some limitations to the use of the social VR task vTime, precisely because of the lack of control of all potent nonverbal language cues, including facial expressions and eye gaze. Therefore, we decided to pursue our efforts to explore visuospatial cognition and nonverbal language during awake surgery by developing an interactive VR task capable of analyzing these functions simultaneously. The possibilities and limitations of this new visuospatial and social VR experience are presented in this paper.

Methods

Study Design

We performed a single-center, prospective, and open-label study, and the study protocol was evaluated and approved by the Agence Nationale de Sécurité du Médicament et des produits de santé, the local ethics committee, and Commission Nationale de l'Informatique et des Libertés. All patients signed a written informed consent form before inclusion in the study. This study was registered at ClinicalTrials.gov (NCT03010943). As indicated above, an amendment was requested and accepted to assess the feasibility of using eye tracking with a VRH during awake craniotomy and to explore the possibilities and limitations of the visuospatial and social VR experience. During the extension of the study, we continued to use questionnaires completed by the patient and medical professionals to assess tolerance (discomfort, nausea, vomiting, and visual-vestibular-somatosensory conflict) and satisfaction (Multimedia Appendix 1). The occurrence of electroencephalogram (EEG) modifications (afterdischarge) or intraoperative seizures (IOSs) was also recorded.

The inclusion criteria were as follows: patients aged >18 years hospitalized for a brain tumor near language and/or motor areas (determined by neuropsychological evaluation and resting-state functional magnetic resonance imaging [fMRI]) in the left or right hemisphere who gave written informed consent. The exclusion criteria were all contraindications for awake surgery (cognitive impairment, whether related to the surgical lesion, aphasia, or morbid anxiety). A total of 15 patients were included in the extension study.

Virtual Reality Headset

This study was performed with a Tobii Pro VR Integration, an eye-tracking retrofitted HTC VIVE wired to a computer connected to a neuronavigational system (Brainlab). The VRH has a visual field of 110°, an adjustable interpupillary distance, a latency <20 milliseconds, a refresh rate of 90 Hz, a resolution of 2160×1200 pixels, and adjustable focus. The VRH includes the eye-tracking systems developed by Tobii Pro for research purposes (Tobii Pro), which collects various types of eye movement data, such as gaze origin and direction, pupil position, and absolute pupil size with an accuracy of 0.5° visual angle at a rate of 120 Hz. What the patient sees in the VRH is visualized on one of the screens of the neuronavigational system.

VR Tasks

The picture-naming task, DO 80, was implemented in the VRH in 2 versions [10]. The first version, in 2D (2D VR), included the same images as the classical naming task DO 80 presented with a computer tablet (an image with the sentence “this is...”). The second version included the same items but in 3D (3D VR), rotating in a virtual, empty space (Figure 1). The object-naming task is simple, making it possible to identify various types of errors, and is the most widely used task for language mapping [12,13].

The new visuospatial and social VR experience that we have developed uses animated synthetic characters (avatars; Figure 2). The scene shows 5 avatars in front of a landscape background. The avatar in the center has his eyes closed and the others in the 4 quadrants of the visualized VRH field are looking in different directions. Patients were asked to search for the avatar staring at them. Less than a second (0.6 seconds) after visual contact is established, the avatar expresses a dynamic facial emotion that the patient is asked to identify and describe: joy, surprise, or anger. If the patient stares at the wrong avatar, a dynamic facial emotion is nevertheless initiated. Facial expressions were enacted by an actor and then transferred onto the avatars with a professional tool from the game or cinema

industry [14]. For each test performed with this VR task, the medical team can follow the gaze of the patient directly, materialized as a green point on one of the screens of the neuronavigational system. This eye tracking reveals whether patients have difficulties exploring the space, have difficulties locating the face of the avatar looking at them, or fail to recognize the facial emotion. At the end of each test, the gaze layout, the time to perform the task, and the answer given by the patient were recorded (Figure 2).

Figure 1. (A) Patient wearing the virtual reality headset. (B) and (C) Example of the item "phone" in the DO 80 naming task presented in 2D (B) and 3D (C) with the virtual reality headset. The green spot indicates the patient's gaze.

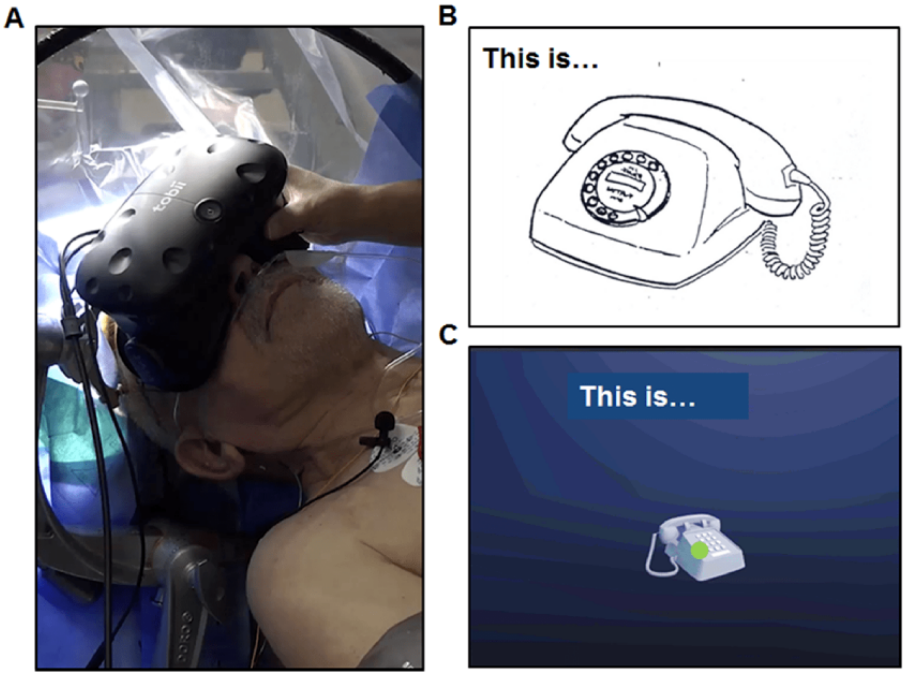
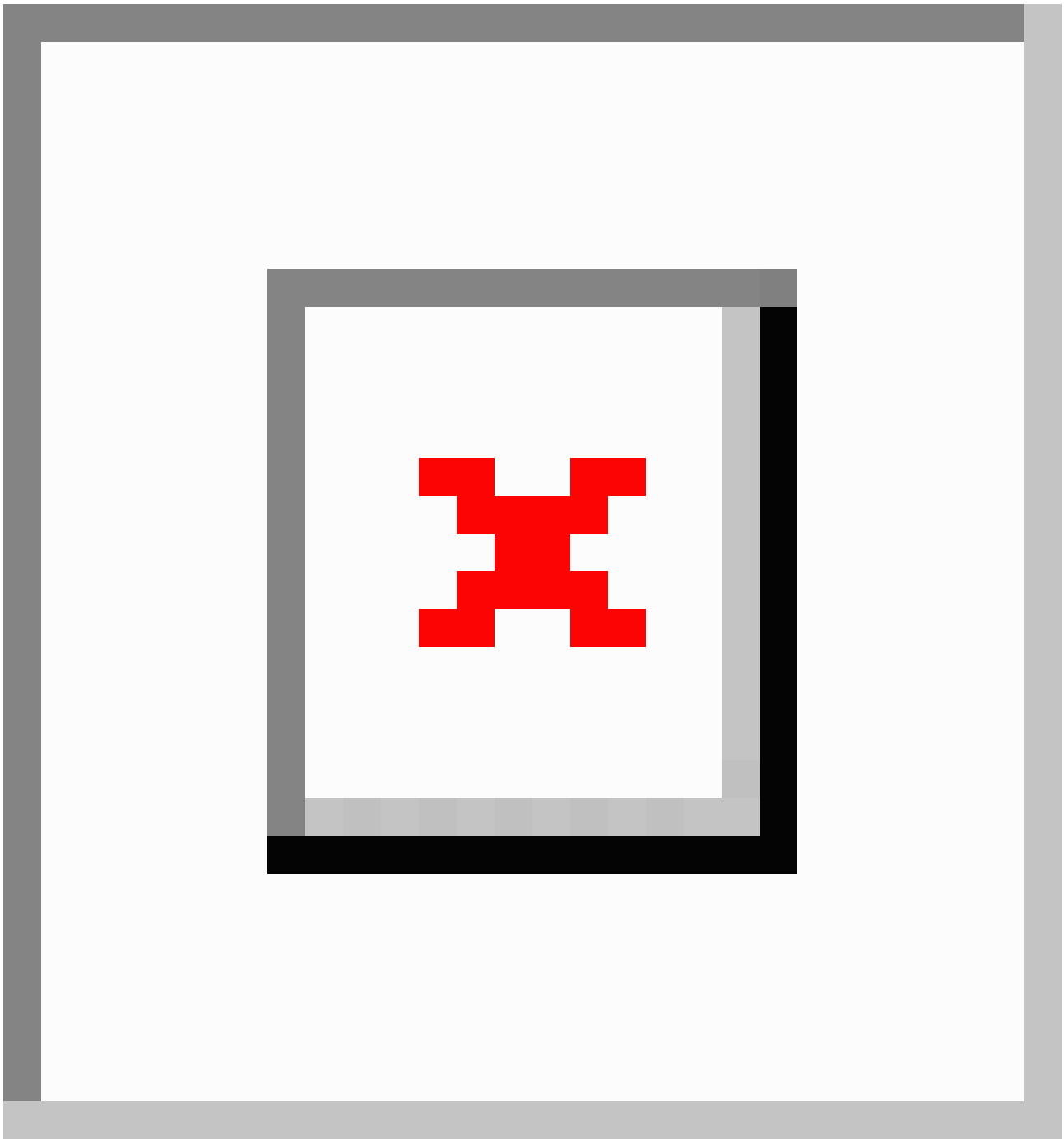


Figure 2. Left: view of the operating room during the procedure. (A) Head of the patient wearing the virtual reality headset; (B) application of direct electrical stimulation to the exposed brain; (C) screen showing what the patient sees in the virtual reality headset, his gaze materialized by a green spot; (D) neuronavigational system showing brain white matter fascicles and the position of the electrode. Right: example of a layout after the virtual reality task simulating a visuospatial and social experience. (E) The image that is visualized and analyzed on the screen (C). The movement of the patient's gaze is visualized as a blue line (with the starting point in green and the endpoint in pink). The green box indicates the avatar making eye contact. The white arrow indicates the avatar on which the patient focuses for more than 0.6 seconds (triggering the expression of a dynamic facial emotion). In this example, the patient identified the avatar making eye contact in 2.53 seconds and indicated the emotion expressed 3.77 seconds later.



Operative Procedure

The procedure has been described in detail elsewhere [9]. Before surgery, all patients underwent a neuropsychological evaluation and fMRI and were trained in the use of VR tasks in the context of awake surgery. General anesthesia was administered via a laryngeal airway mask during surgery. Patients were positioned in a supine or lateral position, according to the location of the tumor, with a rigid pin fixation of the head. The scalp was infiltrated with a local anesthetic. Once the craniotomy, guided by neuronavigation, was completed and the dura had been opened, after infiltration, the patient was woken up. EEG signals were recorded with a four-plot subdural electrode (four-channel Eclipse neurovascular workstation, Medtronic XOMED, Inc). DES was performed with a bipolar electrode (tip-to-tip distance: 5 mm), delivering a biphasic current (parameters: 60 Hz, 1-millisecond single-pulse duration, and current amplitude from 1 mA-8 mA). Stimulation was applied to every 1 cm^2 of the

exposed cerebral cortex. If a functional cortical area was identified, a minimum margin of 1 cm was observed during the resection. We continued monitoring movement and/or spontaneous language during tumor resection, and a second mapping with subcortical electrostimulation was performed if necessary.

Brain mapping for language was performed with the picture-naming task, DO 80, using a computer tablet. Sites were identified as language sites if interference (speech arrest, anomia, dysarthria, semantic or phonemic paraphasia, or delayed naming >5 seconds) was detected in at least 3 meticulous tests (not necessarily consecutively) and were tagged on the cortex. A second round of mapping was then performed using the VRH, with the 2D DO 80 and then with the 3D DO 80 task. Differences in responses were carefully noted. Depending on the location of the tumor, other tests were proposed on a computer tablet (spontaneous speech production, counting, reading, etc).

The visuospatial and social VR experience with avatars was included in brain mapping by DES when considered necessary to test these functions. In other situations, this VR task was proposed for patients without DES, generally during the closure period. The 4 quadrants of the visual field and all emotions (joy, surprise, or anger) were presented randomly to the patient. As for language mapping, a site was identified as eloquent if interference through DES (difficulties exploring the space, difficulties locating the avatar making eye contact, or failure to recognize the facial emotion, resulting in a delay or lack of response from the patient) was detected 3 times. Once the task was completed, the gaze layout, the time to perform the task, and the patient's answer were recorded (Figure 2).

The entire procedure was performed in the presence of an engineer and a neuropsychologist. Heart rate, blood pressure, and EEG signals were recorded continuously during the procedure. Spontaneous or stimulation-induced afterdischarges recorded on EEG, were defined as 2 consecutive spikes or sharp waves distinct from background activity. Any drug administration differing from that laid out in the predefined protocol was noted. Tolerance was also assessed with a questionnaire completed by the patient, the anesthetist, the neuropsychologist, and the neurosurgeon.

Results

Baseline characteristics of the 15 patients are presented in Table 1. A total of 15 patients (8 men and 7 women) with a median age of 52 years (range 25-73 years) underwent the procedure; 2 of the patients were left-handed, and 13 were right-handed. The tumor was in the left hemisphere in 11 patients and the right hemisphere in 4 patients. Patients were initially hospitalized for seizures (9/15, 60%), aphasia (1/15, 7%), or alexia (1/15, 7%). The tumor was discovered through monitoring of the primary cancer in 27% (4/15) of the patients. The mean tumor diameter was 38.6 millimeters (range 25-50 millimeters). Tumors were located in the frontal lobe (8/15, 53%), parietal lobe (4/15, 27%), temporoparietal junction (2/15, 13%), or frontotemporal insular cortex (1/15, 7%). The lesions were grade 2 oligodendroglioma (1/15, 7%), anaplastic oligodendroglioma (2/15, 13%), anaplastic astrocytoma (6/15, 40%), glioblastoma (3/15, 20%), or metastasis (3/15, 20%).

Table 1. Baseline characteristics of the 15 patients and the virtual reality tasks they performed.

Patient	Sex	Age (years)	Handedness	Diagnosis	Hemisphere	Lobe	Preoperative training	Brain mapping
1	Male	68	Left	Metastasis	Left	Parietal	Task 1[a] and task 2[b]	Task 1 and task 2
2	Male	41	Right	Oligodendroglioma II	Right	Frontal	Task 1 and task 2	Motor and task 2
3	Female	25	Right	Astrocytoma III	Left	Frontal	Task 1	Motor and task 1
4	Female	66	Right	Oligodendroglioma III	Right	Frontal	Task 2	Motor
5	Male	39	Left	Astrocytoma III	Right	Frontal	Task 1 and task 2	Motor and task 1 and task 2
6	Female	60	Right	Glioblastoma	Left	Temporoparietal	Task 1	Task 1
7	Male	48	Right	Oligodendroglioma III	Left	Frontal	Task 1 and task 2	Task 1 and task 2
8	Female	53	Right	Glioblastoma	Left	Parietal	Task 1	Task 1
9	Male	68	Right	Glioblastoma	Left	Frontal	Task 1 and task 2	Task 1
10	Male	73	Right	Metastasis	Left	Frontal	Task 1 and task 2	Task 1 and task 2
11	Female	47	Right	Astrocytoma III	Left	Parietal	Task 1 and task 2	Motor and task 1
12	Male	61	Right	Metastasis	Left	Temporoparietal	Task 1 and task 2	Task 1
13	Male	43	Right	Astrocytoma III	Left	Parietal	Task 1 and task 2	Task 1 and task 2
14	Female	53	Right	Astrocytoma III	Left	Frontotemporal insular	Task 1 and task 2	Task 1
15	Female	41	Right	Astrocytoma III	Right	Frontal	Task 2	Task 2

[a]Task 1: DO 80 (tablet, 2D virtual reality, and 3D virtual reality).

[b]Task 2: visuospatial and social virtual reality experience.

Only 3 patients had experienced VR before inclusion. Before surgery, 13 patients were trained with the DO 80 task (tablet, 2D VR, and 3D VR), and 12 patients were trained with the VR task simulating a visuospatial and social experience (Table 1). No preoperative difficulties were observed with the DO 80 task, but 2 patients (patients 4 and 12) experienced difficulties with visuospatial and social VR tasks. In particular, they encountered problems in finding the avatar looking at them. For these 2 patients, this task was not applied during awake surgery (Table 1).

The mean duration of surgery was 4 hours and 23 minutes (range 3 h and 6 min-5 h and 30 min), with a mean duration of the awake phase of 2 hours and 20 minutes (range 25 min-4 h). The mean intensity of DES was 1.9 mA (range 1-4 mA), and the mean total duration of VRH use per patient was 11 minutes in 2 to 4 sessions.

For the 13 patients for whom brain mapping was performed for language, the same language eloquent areas were identified, regardless of the DO 80 presentation used (computer tablet, 2D VR, or 3D VR). However, for 1 patient (patient 13), the results were unclear in some areas for DO 80 on the computer tablet (hesitation or delay in denomination) that clearly were not eloquent according to assessment with the VRH. Eye tracking was functional, making it possible to trace the gaze of the patient during the task. During the DO 80 task, we noted that patients did not read the sentence “this is...,” instead saying it automatically.

Among the 10 patients who were able to perform the visuospatial and social VR experience without difficulty before surgery, 7 patients performed this task during brain mapping by DES and 2 patients (patients 11 and 14) during closure without DES (Table 1). For 1 patient (patient 9), it was not possible to present the task at the end of surgery because of hemostasis problems. Without DES, the mean time taken to identify eye contact was 2.3 seconds (range 2.0-2.5 seconds), and the mean time taken to recognize and verbalize the facial emotion was 3.2 seconds (range 2.6-4.4 seconds; total time for the task: 5.5 seconds, range 4.9-7.0 seconds). This total time is, in reality, shorter because the test is stopped manually once the patient’s response has been heard. For 2 patients (patients 2 and 15), DES of an area in the right hemisphere disturbed visual exploration and delayed avatar identification.

Despite the discomfort associated with the awake surgery procedure, none of the patients experienced vertigo or any vegetative signs of VR sickness. EEG modifications (afterdischarge or spike-and-wave) were observed in 27% (4/15) of the patients during the standard brain mapping procedure (without VRH). The same abnormalities persisted during brain

mapping+VRH in 3 of these patients. IOSs occurred in 13% (2/15) of the patients. Epilepsy was the first sign for these patients and neither of these two patients displayed EEG modifications during the brain mapping procedure. The IOSs observed were short motor seizures, disappearing rapidly after cortical irrigation with iced saline. IOSs occurred during DES, before using the VRH for one patient and during the VR task for the other.

According to the questionnaire completed after surgery by the patient, the neurosurgeon, and the anesthetist, the use of the VRH was not an issue during surgery. During 1 operation, the neuropsychologist found it difficult to position the VRH. All participants agreed to continue studying this approach.

Discussion

Principal Findings

VR is a domain with growing applications in the field of neuroscience. This computer technology generates realistic images, sounds, and other sensations that simulate a user's physical presence in a virtual or imaginary environment. A person using a VRH can look around the artificial world, *move* within it, and interact with virtual features or items. As such, VRH provides a unique opportunity to combine the naturalness of everyday interactions with the experimental controls required during brain mapping procedures, paving the way for new brain mapping procedures for complex cognitive functions.

The extension of our initial prospective trial confirmed that VRH with eye tracking and immersive virtual experiences was safe for patients undergoing awake craniotomy and brain mapping using DES. None of the patients experienced VR sickness and we observed no sympathetic nervous activity reported for this syndrome [15-20]. On the basis of our personal experience and published data, we are convinced that this good tolerance was because of patient preparation and training [21-23]. In total, in the 2 studies (45 patients), we observed afterdischarges in 17 (38%) patients and IOSs in 11 (24%) patients, rates within the range reported in previous studies: 71% for afterdischarges [24] and 3.4% to 31% for IOSs [25-29]. The IOS rate cannot be explained by the use of VRH, as most seizures occurred before its implementation. Thus, these seizures are more likely to be owing to our brain mapping procedure, which always began with positive motor stimulation to calibrate DES intensity. Consistent with our hypothesis, a recent review showed that patients with positive mapping (detection of the functional cortical area) are at a higher risk of IOSs [30]. Patients with preoperative seizures were also found to be more susceptible to intraoperative or postoperative seizures; 8 of the 11 patients with intraoperative IOSs in our 2 studies fell into this category. Another explanation would be the use of perioperative EEG: there is some controversy concerning the effect of brain activity monitoring on the occurrence of IOSs. Nossek et al [31] compared the occurrence of IOSs in the presence and absence of electrocorticogram (ECoG) use and found that the use of ECoG was associated with an increase in the occurrence of IOSs. The authors suggested that this was probably because the surgeon tends to increase stimulation more liberally when ECoG is used than when it is not.

All these data indicate that the use of a VRH during brain mapping in awake surgery does not specifically increase the rate of IOSs. Nevertheless, we recommend several precautions to prevent seizures during the use of a VRH for brain mapping procedures, including a well-trained team and, although there is no consensus regarding its usefulness, intraoperative monitoring of brain electrical activity.

This trial also demonstrated the feasibility of using eye tracking in patients undergoing awake craniotomy and brain mapping using DES. One of our apprehensions was the potential interference caused by devices in the operating room emitting infrared light, such as the neuronavigation system for tracking gaze. No such interference was observed and we were able to track eye movements on one of the screens of the neuronavigational system. Eye tracking revealed that the patients never read the sentence "this is..." associated with the image in the DO 80 VR task. This observation suggests that the patient said the sentence automatically, focusing only on the naming task. In the future, it would be interesting to develop a new VR task through a VRH with eye tracking for the specific exploration of reading. We recognize that eye-tracking sensors can be used in combination with regular computer screens or tablets. However, the eye tracker in the VRH, which combines features of both mobile and remote setups, prevents the risk of losing calibration and improves the success rate for measurements of eye positions and movements. Furthermore, the use of a VRH immerses the patient in the VR task, which is completely isolated from the surrounding operating room.

This trial also explored the possibilities and limitations of the visuospatial and social VR experiment developed with animated synthetic avatars. Avatars are perceived in a similar manner to real human beings and can be used to explore the complex processes of nonverbal language, empathy, and theory of mind [32]. Moreover, avatars enable researchers to manipulate, in a selective manner, variables that cannot be independently investigated in naturalistic situations, allowing precise control of not only the intensity, kinetics, and type of emotion but also of facial physiognomy, race, sex, and age as a function of the paradigm used. Tasks for use in awake surgery must take about 5 seconds, corresponding to the duration of DES. During this period, the maximum number of faces that can be analyzed in a visual field of 110° is limited, even if faces have the spatial advantage of capturing attention, reflecting their particular saliency, as well as their social value. Therefore, we decided to include 5 avatars, 1 in the center and the others in the 4 quadrants of the visualized field, to allow sufficient spatial exploration. The simulated social interaction involves the patients making visual contact with an avatar looking at them and describing the automatically triggered facial emotion of the avatar or their feelings about the desire for communication or social contact expressed by the avatar. Direct gazes between 2 people are known to constitute a significant, engaging social signal in all cultures. Decoding the movement of gaze plays an important role in predicting intentions and can be regarded as an important element in the theory of mind [33-36]. The patient's eye-tracking data were used for real-time control of the virtual character's emotional behavior in response to the participant's gaze. We showed that this VR experience was

compatible with the brain mapping procedure, with answers obtained within the maximum time allowable for DES. We observed that DES in some areas of the brain disturbed the answer to the test. Through eye tracking and patients' answers, it was possible to attribute these failures to difficulties in exploring the space, identifying eye contact, or recognizing the facial emotion expressed or the associated mental state. A prospective study (ClinicalTrials.gov NCT04288505) is currently underway to determine the specificity and sensitivity of this new visuospatial and social VR task before its introduction into routine use for exploring the neural substrates of visuospatial and social functions during awake brain surgery. Several established pencil-and-paper tests are being used to determine its performance, including the bells test [37] for visuospatial attention functions and the Ekman test [38] and the *Reading the Mind in the Eyes* test [39] for social cognitive functions.

Limitations

One of the difficulties in the field of VR research is the rapid progress of technology and the regular release of new VRHs. At the beginning of our research on the use of VR in the operating room, intending to detect hemianopsia and unilateral neglect during DES, we used the Oculus VRHs DK1 and DK2 (visual field 100°, resolution 1280×800 pixels, and refresh rate 60 Hz; Oculus) [4]. For the tolerance study, we chose to use a wireless low-cost, high-quality, customizable device: the Samsung Gear VR combined with a Samsung S7 smartphone (Android platform; visual field 96°, resolution 1440×1280 pixels, and refresh rate 60 Hz) [9]. However, before the completion of this study, new VRH models, including an eye-tracking system, became available. We chose to prolong the study with a higher-performance VRH, the HTC VIVE (larger visual field 110°, better resolution 2160×1200 pixels, and higher refresh rate 90 Hz) combined with an eye-tracking device (Tobii Pro SDK) capable of tracking the full HTC VIVE field of view and measuring the pupil.

Conclusions

This study extended an initial prospective trial designed to confirm the feasibility and safety of VRH use and immersive virtual experiences for patients undergoing awake craniotomy and brain mapping using DES. Its added value lies in using a latest generation VRH, including eye tracking and a VR task designed to simultaneously test visuospatial and social functions.

Acknowledgments

This project received funding from Fondation de l'Avenir, Paris, France (AP-RM 18-032), and Angers University Hospital. The authors thank Doctor Jean-Michel Lemée, Doctor Florian Bernard (Département de Neurochirugie, Centre Hospitalier Universitaire d'Angers, Angers, France), Doctor Jérémy Besnard, and Professor Philippe Allain (LPPL-EA4638, Université d'Angers, Angers, France) for their help in the design of the new clinical trial (ClinicalTrials.gov NCT04288505). The authors also thank Alex Edelman and Associates (La Crouzille, France) for correcting the manuscript.

Conflicts of Interest

RS is a cofounder of Dynamixyz, which markets the facial expression transfer tool used to animate avatars. He reports personal fees from Dynamixyz. None of the other authors have any conflicts of interest to declare.

Multimedia Appendix 1
Questionnaires completed by the patients and medical professionals.
[DOC File , 154 KB - jmir_v23i3e24373_app1.doc]

References

1. Freyschlag CF, Duffau H. Awake brain mapping of cortex and subcortical pathways in brain tumor surgery. J Neurosurg Sci 2014 Dec;58(4):199-213. [Medline: 25418274]
2. Sanai N, Mirzadeh Z, Berger MS. Functional outcome after language mapping for glioma resection. N Engl J Med 2008 Jan 03;358(1):18-27. [doi: 10.1056/NEJMoa067819] [Medline: 18172171]
3. Bernard F, Lemée JM, Ter Minassian A, Menei P. Right Hemisphere Cognitive Functions: From Clinical and Anatomic Bases to Brain Mapping During Awake Craniotomy Part I: Clinical and Functional Anatomy. World Neurosurg 2018 Oct;118:348-359. [doi: 10.1016/j.wneu.2018.05.024] [Medline: 29763748]
4. Mazerand E, Le RM, Hue S, Lemée JM, Klinger E, Menei P. Intraoperative Subcortical Electrical Mapping of the Optic Tract in Awake Surgery Using a Virtual Reality Headset. World Neurosurg 2017 Jan;97:424-430. [doi: 10.1016/j.wneu.2016.10.031] [Medline: 27756671]
5. Kasteleijn-Nolst Trenité DGA, Martins da Silva A, Ricci S, Rubboli G, Tassinari CA, Lopes J, et al. Video games are exciting: a European study of video game-induced seizures and epilepsy. Epileptic Disord 2002 Jun;4(2):121-128 [FREE Full text] [Medline: 12105074]
6. Merhi O, Faugloire E, Flanagan M, Stoffregen TA. Motion sickness, console video games, and head-mounted displays. Hum Factors 2007 Oct;49(5):920-934. [doi: 10.1518/001872007X230262] [Medline: 17915607]

7. Sharples S, Cobb S, Moody A, Wilson JR. Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. Displays 2008 Mar;29(2):58-69. [doi: 10.1016/j.displa.2007.09.005]
8. Shoja MM, Tubbs RS, Malekian A, Jafari Rouhi AH, Barzgar M, Oakes WJ. Video game epilepsy in the twentieth century: a review. Childs Nerv Syst 2007 Mar;23(3):265-267. [doi: 10.1007/s00381-006-0285-2] [Medline: 17211654]
9. Delion M, Klinger E, Bernard F, Aubin G, Ter Minassian A, Menei P. Immersing Patients in a Virtual Reality Environment for Brain Mapping During Awake Surgery: Safety Study. World Neurosurg 2020 Feb;134:e937-e943. [doi: 10.1016/j.wneu.2019.11.047] [Medline: 31734424]
10. Deloche G, Hannequin D. Test de dénomination orale d'images DO 80. Paris: Les Editions du Centre de Psychologie Appliquée; 1997.
11. Bernard F, Lemée JM, Aubin G, Ter Minassian A, Menei P. Using a Virtual Reality Social Network During Awake Craniotomy to Map Social Cognition: Prospective Trial. J Med Internet Res 2018 Jun 26;20(6):e10332 [FREE Full text] [doi: 10.2196/10332] [Medline: 29945859]
12. Duffau H, Capelle L, Denvil D, Sichez N, Gatignol P, Taillandier L, et al. Usefulness of intraoperative electrical subcortical mapping during surgery for low-grade gliomas located within eloquent brain regions: functional results in a consecutive series of 103 patients. J Neurosurg 2003 Apr;98(4):764-778. [doi: 10.3171/jns.2003.98.4.0764] [Medline: 12691401]
13. Ojemann JG, Ojemann GA, Lettich E. Cortical stimulation mapping of language cortex by using a verb generation task: effects of learning and comparison to mapping based on object naming. J Neurosurg 2002 Jul;97(1):33-38. [doi: 10.3171/jns.2002.97.1.0033] [Medline: 12134930]
14. Dynamixyz. URL: https://www.dynamixyz.com/ [accessed 2021-03-12]
15. Buker TJ, Vincenzi DA, Deaton JE. The effect of apparent latency on simulator sickness while using a see-through helmet-mounted display: reducing apparent latency with predictive compensation. Hum Factors 2012 Apr;54(2):235-249. [doi: 10.1177/0018720811428734] [Medline: 22624290]
16. Chen W, Chao JG, Chen XW, Wang JK, Tan C. Quantitative orientation preference and susceptibility to space motion sickness simulated in a virtual reality environment. Brain Res Bull 2015 Apr;113:17-26. [doi: 10.1016/j.brainresbull.2015.01.007] [Medline: 25647233]
17. Kim YY, Kim HJ, Kim EN, Ko HD, Kim HT. Characteristic changes in the physiological components of cybersickness. Psychophysiology 2005 Sep;42(5):616-625. [doi: 10.1111/j.1469-8986.2005.00349.x] [Medline: 16176385]
18. Malińska M, Zużewicz K, Bugajska J, Grabowski A. Heart rate variability (HRV) during virtual reality immersion. Int J Occup Saf Ergon 2015;21(1):47-54 [FREE Full text] [doi: 10.1080/10803548.2015.1017964] [Medline: 26327262]
19. Nalivaiko E, Davis SL, Blackmore KL, Vakulin A, Nesbitt KV. Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. Physiol Behav 2015 Nov 01;151:583-590. [doi: 10.1016/j.physbeh.2015.08.043] [Medline: 26340855]
20. Ohyama S, Nishiike S, Watanabe H, Matsuoka K, Akizuki H, Takeda N, et al. Autonomic responses during motion sickness induced by virtual reality. Auris Nasus Larynx 2007 Sep;34(3):303-306. [doi: 10.1016/j.anl.2007.01.002] [Medline: 17336009]
21. Ghazanwy M, Chakrabarti R, Tewari A, Sinha A. Awake craniotomy: A qualitative review and future challenges. Saudi J Anaesth 2014 Oct;8(4):529-539 [FREE Full text] [doi: 10.4103/1658-354X.140890] [Medline: 25422613]
22. Kelm A, Sollmann N, Ille S, Meyer B, Ringel F, Krieg SM. Resection of Gliomas with and without Neuropsychological Support during Awake Craniotomy-Effects on Surgery and Clinical Outcome. Front Oncol 2017;7:176 [FREE Full text] [doi: 10.3389/fonc.2017.00176] [Medline: 28868255]
23. Pallud J, Rigaux-Viode O, Corns R, Muto J, Lopez Lopez C, Mellerio C, et al. Direct electrical bipolar electrostimulation for functional cortical and subcortical cerebral mapping in awake craniotomy. Practical considerations. Neurochirurgie 2017 Jun;63(3):164-174. [doi: 10.1016/j.neuchi.2016.08.009] [Medline: 28162255]
24. Pouratian N, Cannestra AF, Bookheimer SY, Martin NA, Toga AW. Variability of intraoperative electrocortical stimulation mapping parameters across and within individuals. J Neurosurg 2004 Sep;101(3):458-466. [doi: 10.3171/jns.2004.101.3.0458] [Medline: 15352604]
25. Conte V, Carrabba G, Magni L, L'Acqua C, Magnoni S, Bello L, et al. Risk of perioperative seizures in patients undergoing craniotomy with intraoperative brain mapping. Minerva Anestesiol 2015 Apr;81(4):379-388. [Medline: 25057931]
26. Eseonu CI, Rincon-Torroella J, Lee YM, ReFaey K, Tripathi P, Quinones-Hinojosa A. Intraoperative Seizures in Awake Craniotomy for Perirolandic Glioma Resections That Undergo Cortical Mapping. J Neurol Surg A Cent Eur Neurosurg 2018 May;79(3):239-246. [doi: 10.1055/s-0037-1617759] [Medline: 29346829]
27. Nossek E, Matot I, Shahar T, Barzilai O, Rapoport Y, Gonen T, et al. Intraoperative seizures during awake craniotomy: incidence and consequences: analysis of 477 patients. Neurosurgery 2013 Jul;73(1):135-40; discussion 140. [doi: 10.1227/01.neu.0000429847.91707.97] [Medline: 23615101]
28. Pouratian N, Cannestra AF, Bookheimer SY, Martin NA, Toga AW. Variability of intraoperative electrocortical stimulation mapping parameters across and within individuals. J Neurosurg 2004 Sep;101(3):458-466. [doi: 10.3171/jns.2004.101.3.0458] [Medline: 15352604]

29. Wang YC, Lee CC, Takami H, Shen S, Chen KT, Wei KC, et al. Awake craniotomies for epileptic gliomas: intraoperative and postoperative seizure control and prognostic factors. J Neurooncol 2019 May;142(3):577-586. [doi: 10.1007/s11060-019-03131-0] [Medline: 30805752]
30. Roca E, Pallud J, Guerrini F, Panciani PP, Fontanella M, Spena G. Stimulation-related intraoperative seizures during awake surgery: a review of available evidences. Neurosurg Rev 2020 Feb;43(1):87-93. [doi: 10.1007/s10143-019-01214-0] [Medline: 31797239]
31. Nossek E, Matot I, Shahar T, Barzilai O, Rapoport Y, Gonen T, et al. Failed awake craniotomy: a retrospective analysis in 424 patients undergoing craniotomy for brain tumor. J Neurosurg 2013 Feb;118(2):243-249. [doi: 10.3171/2012.10.JNS12511] [Medline: 23121432]
32. de Borst AW, de Gelder B. Is it the real deal? Perception of virtual characters versus humans: an affective cognitive neuroscience perspective. Front Psychol 2015;6:576 [FREE Full text] [doi: 10.3389/fpsyg.2015.00576] [Medline: 26029133]
33. Byom LJ, Mutlu B. Theory of mind: mechanisms, methods, and new directions. Front Hum Neurosci 2013;7:413 [FREE Full text] [doi: 10.3389/fnhum.2013.00413] [Medline: 23964218]
34. Shepherd SV. Following gaze: gaze-following behavior as a window into social cognition. Front Integr Neurosci 2010;4:5 [FREE Full text] [doi: 10.3389/fnint.2010.00005] [Medline: 20428494]
35. Hamilton AFC. Gazing at me: the importance of social meaning in understanding direct-gaze cues. Philos Trans R Soc Lond B Biol Sci 2016 Jan 19;371(1686):20150080 [FREE Full text] [doi: 10.1098/rstb.2015.0080] [Medline: 26644598]
36. Cañigueral R, Hamilton AFC. The Role of Eye Gaze During Natural Social Interactions in Typical and Autistic People. Front Psychol 2019;10:560 [FREE Full text] [doi: 10.3389/fpsyg.2019.00560] [Medline: 30930822]
37. Ferber S, Karnath HO. How to assess spatial neglect--line bisection or cancellation tasks? J Clin Exp Neuropsychol 2001 Oct;23(5):599-607. [doi: 10.1076/jcen.23.5.599.1243] [Medline: 11778637]
38. Ekman P, Friesen WV. Measuring facial movement. J Nonverbal Behav 1976;1(1):56-75. [doi: 10.1007/bf01115465]
39. Baron-Cohen S, Bowen DC, Holt RJ, Allison C, Auyeung B, Lombardo MV, et al. The "Reading the Mind in the Eyes" Test: Complete Absence of Typical Sex Difference in ~400 Men and Women with Autism. PLoS ONE 2015 Aug 27;10(8):e0136521. [doi: 10.1371/journal.pone.0136521]

Abbreviations

DES: direct electrical stimulation
ECoG: electrocorticogram
EEG: electroencephalogram
fMRI: functional magnetic resonance imaging
IOS: intraoperative seizure
VR: virtual reality
VRH: virtual reality headset

Edited by G Eysenbach; submitted 16.09.20; peer-reviewed by D Lin, M Martelli; comments to author 05.01.21; revised version received 26.01.21; accepted 16.02.21; published 24.03.21.

Please cite as:
Casanova M, Clavreul A, Soulard G, Delion M, Aubin G, Ter Minassian A, Seguier R, Menei P
Immersive Virtual Reality and Ocular Tracking for Brain Mapping During Awake Surgery: Prospective Evaluation Study
J Med Internet Res 2021;23(3):e24373
URL: https://www.jmir.org/2021/3/e24373
doi:10.2196/24373
PMID:33759794

Original Paper

Classification of the Use of Online Health Information Channels and Variation in Motivations for Channel Selection: Cross-sectional Survey

Di Zhang[1], PhD; Zhen Shi[1], MSc; Hongchao Hu[1], MA; Gang (Kevin) Han[2], PhD

[1]The Research Center of Journalism and Social Development, Renmin University of China, Beijing, China

[2]Greenlee School Journalism and Communication, Iowa State University, Ames, IA, United States

Corresponding Author:
Gang (Kevin) Han, PhD
Greenlee School Journalism and Communication
Iowa State University
119 Hamilton
613 Wallace Rd
Ames, IA, 50011-4010
United States
Phone: 1 515 294 0482
Email: ghan@iastate.edu

Abstract

Background: Existing health education and communication research routinely measures online channel use as a whole by, for example, evaluating how frequently people use the internet to search for health information. This approach fails to capture the complexity and diversity of online channel use in health information seeking. The measurement of generic online channel use may cause too much error, and it lends no support to media planning in public health promotion campaigns or scholarly research involving online channel use.

Objective: This study intends to present a thorough picture of patterns of online health information channel use and classify the use of various types of online health information channels, including WeChat, microblogs, web portals, search engines, mobile apps, and online forums. Under the framework of the risk information seeking and processing model, this study also analyzes the differences in individuals' motivations for channel selection to offer further evidence to validate the classification scheme.

Methods: This study sampled 542 Chinese internet users in Beijing. The average age of the respondents was 33 years, female respondents accounted for 52.0% (282/542) of the sample, and the average monthly income ranged from US $900 to $1200. The study surveyed the use of 13 commonly used online health information channels and various sociopsychological factors associated with online health information seeking.

Results: This study derived 3 categories of online health information channels: searching, browsing, and scanning channels. It was found that the use of online searching channels was affect driven (B=0.11; β=0.10; *P*=.02) and characterized by a stronger need for health knowledge (B=0.09; β=0.01; *P*<.001). The use of browsing channels was directly influenced by informational subjective norms (B=0.33; β=0.15; *P*=.004) and perceived current knowledge (B=0.007; β=0.09; *P*=.003). The use of scanning channels was mainly influenced by informational subjective norms (B=0.29; β=0.15; *P*=.007).

Conclusions: The results of this study suggest that health communication practitioners and scholars may consider measuring the use of internet, new media, or online media more precisely instead of simply asking the public about the frequency of online channel use or internet use in the acquisition of health information. Scholars and practitioners may consider measuring the use of online health information channels by using the 3-category scheme described in this study. Future research is encouraged to further explore how people process health information when using different online channels.

(J Med Internet Res 2021;23(3):e24945) doi:10.2196/24945

KEYWORDS

search; browse; scan; health information seeking; channel selection; health information; health education; health communication; online media

Introduction

Background

People across the globe increasingly name online channels as one of their top choices when acquiring health information and knowledge [1], which accordingly attracts scholarly attention. Most existing studies treat online channels as a whole by, for example, evaluating how frequently people use the internet to search for health information [2,3]. In this study, online health information channels refer to online communication media and applications that collect health information and knowledge from sources, repackage them, and then distribute them to people [4]. However, terms such as "new media," "online media," and "internet" are very generic. Treating the internet as a whole does not reflect the complexity and diversity in the use of online health information channels [5]. Therefore, the first goal of this study is to survey the use of various types of online health information channels, thus presenting a more nuanced picture of the use of online health information channels.

In addition to describing the channel use frequencies, the second goal is to classify the use of various online health information channels into different categories from the perspective of information-seeking behaviors, which differs from the existing division method. When there is a need to analyze online channels separately, health communication researchers routinely divide online health information channels into news portals (such as news websites and health websites) and social media, as in the case of the Health Information National Trends Survey. This classification scheme is based on the differences in the diversity of content creators, platform structure, and connections between users [6], which is inherently the perspective of the platform designers and operators rather than the users. However, this perspective is increasingly incompatible with consumer-centered health communication campaign design, which assumes that satisfying individual needs is the key to effective campaigns [7]. In the era of the internet, it is more noticeable that individuals select communication channels to meet their felt needs [8]. Given that individual needs drive different styles of information-seeking behaviors, such as searching, browsing, and scanning [9,10], the development of an information-seeking, behavior-based channel classification scheme can be potentially more helpful for contemporary health communication researchers and practitioners, who would be able to use fewer measurement items while increasing the validity of the measurement in both academic and formative research of health information channel use. The development of such a scheme is also feasible because different channel types can fulfill different information-seeking strategies [11].

Third, this study examines the factors associated with online health information channel selection under the framework of the risk information seeking and processing (RISP) model, which depicts the various sociopsychological factors behind information seeking and processing [12]. The results of the study can contribute to the literature on RISP by explaining the variance in channel selection, which is a crucial part of information seeking that is underexplored [13].

Literature Review

Classification of Online Channel Use and Health Information–Seeking Behaviors

With the rise of the internet and mobile phones, people in countries like the United States and China frequently search for health information on the internet [14,15]. Additionally, evidence collected in multiple countries revealed that people who search for health information use various types of online channels, such as search engines, health web portals, social networking sites, and online support groups [14,16,17].

This study intended to pinpoint the underlying patterns by classifying the use of online health information channels into clusters based on health information acquisition behaviors, which differs from the existing division method from the perspective of platform designers and operators. Current research on health information seeking suggests 3 types of behaviors: searching, browsing, and scanning [9,10]. People acquire health information mainly through 2 routes, active seeking and scanning [9]. Active seeking refers to an intentional process of acquiring health information, which implies more active efforts in health information seeking [9,18]. Research in library and information science and health informatics implies that active seeking can be further divided into searching and browsing, which vary in the degree of specificity of information seeking [19]. In health information, searching, which is directed, refers to users searching the internet for answers related to specific diseases or symptoms [10,20]. Browsing, which is undirected and motivated by curiosity, refers to users browsing health information without the intent to acquire knowledge about specific diseases or symptoms, consuming health information regularly and habitually, and following the structure and layout of information prepared by the publisher of the website or account [10,20]. In contrast, scanning is defined as a process in which people both encounter health information or knowledge while engaged in tasks unrelated to health and make a decision to process it [9], which is an effort less active than active seeking [21].

Different channel types can fulfill different information-seeking strategies because of the interface designs and functionalities of these channels [11]. According to interviews before data collection, people generally use a specific online channel to engage in the primary information acquisition activity. However, one particular channel can be suitable for more than one type of health information acquisition behavior. For instance, web portals can apply to both browsing and scanning, so web portals fall into both categories in this study. From what has been discussed above, the following research question was proposed: What are the patterns of online health information channel use?

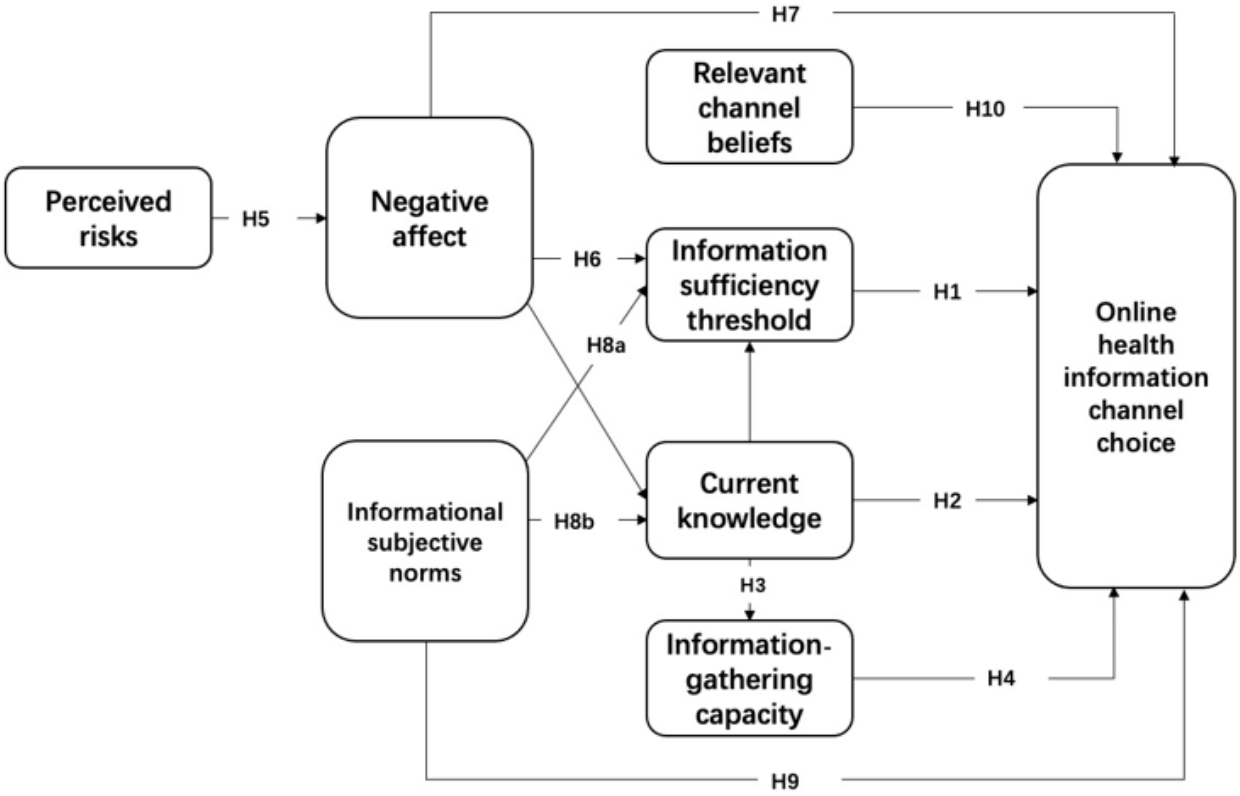
Health Information Channel Selection and RISP

As opposed to many existing channel choice studies that emphasize the influences of channel characteristics (eg, ease of use, interactivity, privacy, and media scale) [22,23], this study explains the channel choice using an audience-centered approach because channel use depends on audience needs, particularly their psychological needs [4,24,25]. Specifically, the study used RISP as the theoretical model, which combines theories such

as the heuristic-systematic model and the theory of planned behavior [12,26-29]. RISP was chosen for two other reasons. First, RISP has often been applied in the context of health risks, such as environmental health [30], vaccines [31], and clinical trial enrollment [32]. Second, channel choice can reflect different combinations of information-seeking and processing strategies, two major dependent variables in RISP. Different types of online health information channel use vary in information-seeking strategies (routine [nonactive] vs nonroutine [active]) [11,33] and may result in differences in information processing (heuristic vs systematic) [12]. In RISP, a combination of the 4 attributes in information seeking and processing results in a fourfold typology: (1) routine (ie, habitual, ritual) seeking with heuristic processing, (2) routine seeking with systematic processing, (3) nonroutine seeking with heuristic processing, and (4) nonroutine seeking with systematic processing [12]. In reference to the abovementioned conceptual definitions of browsing, searching, and scanning, channel choice based on information-seeking strategies may represent different combinations of information-seeking and processing strategies, thus making RISP an ideal theoretical framework. For instance, searching (active but directed) is similar to nonroutine seeking with systematic processing.

RISP studies primarily focus on the following key predictor variables: information insufficiency, informational subjective norms, perceived hazard characteristics, affective responses, relevant channel beliefs, and perceived information-gathering capacities. Figure 1 shows the conceptual model consisting of predictor and outcome variables.

Figure 1. Conceptual model. The lines between current knowledge and negative affect and information sufficiency threshold were added for the purpose of statistical control. Thus, these lines are unmarked. H: hypothesis.



Information insufficiency refers to the perceived cognitive need for additional information, which is the perception of a gap between one's existing knowledge and a level of knowledge sufficient to handle risks confidently (information sufficiency threshold) [12,33]. Previous studies confirm that information insufficiency leads to active information seeking and systematic processing [34,35]. Additionally, perceived knowledge is essentially the complement to the sufficiency threshold link for purposes of assessing regressed change and difference scores; RISP meta-analyses and marketing research suggest that current knowledge accounts for a substantial amount of variance in information seeking [36] and systematic processing [37-39]. Since previous RISP studies did not directly examine channel choice, 2 sets of hypotheses without directions were proposed for the 2 variables:

- H1: Controlling for current knowledge, the information sufficiency threshold is related to the use of searching sites (H1a), browsing sites (H1b), and scanning sites (H1c).
- H2: Current knowledge is related to the use of searching sites (H2a), browsing sites (H2b), and scanning sites (H2c).

Perceived information-gathering capacity refers to one's perceived abilities to acquire and process risk-related information [26,29,33], which is positively associated with current knowledge and information seeking [30,40]. Motivated by information insufficiency or pressure from social norms to seek information, people may need confidence that they are able to perform information-seeking tasks [41]. Thus, the following were proposed:

- H3: Current knowledge is positively associated with perceived information-gathering capacity.

- H4: Perceived information-gathering capacity is related to the use of searching sites (H4a), browsing sites (H4b), and scanning sites (H4c).

Perceived hazard characteristics are one's cognitive evaluation of the nature of hazards [33]. Perceived hazard characteristics include personal control, trust in risk management, perceived threats to personal values, and risk judgment [12,33]. However, this study focused on the perceived probability of contracting a disease (short for perceived risk afterwards) because most people are unlikely to have serious diseases. Previous RISP studies have found that risk judgment can potentially increase the level of negative affect [12,41,42], which further contributes to more information seeking and processing via one's information insufficiency level or directly [40]. Thus, the following was proposed:

- H5: Perceived risks are positively associated with negative affect.

Affective responses are induced by risk perception and can lead to a higher level of information insufficiency [12]. Negative affect was found to influence information-seeking behaviors directly [40]. Thus, the following were proposed:

- H6: Negative affect is positively associated with the information sufficiency threshold.
- H7: Negative affect is related to the use of searching sites (H7a), browsing sites (H7b), and scanning sites (H7c).

Informational subjective norms are the perceived socioenvironmental influence on an individual's subjective assessment of information held to handle a given risk and motivation to seek and process information [33]. Informational subjective norms consist of two dimensions, injunctive and descriptive [35], which have been found to influence information seeking [36,43]. Thus, the following were proposed:

- H8: Informational subjective norms are positively associated with the information sufficiency threshold (H8a) and current knowledge (H8b).
- H9: Informational subjective norms are related to the use of searching sites (H9a), browsing sites (H9b), and scanning sites (H9c).

Relevant channel beliefs refer to people's beliefs about channels that carry risk-related information, including their trustworthiness and usefulness, that could influence information seeking and processing either as a main effect or as a moderator [12,33]. Thus, the following was proposed:

- H10: Relevant channel beliefs are related to the use of searching sites (H10a), browsing sites (H10b), and scanning sites (H10c).

Methods

Sample

This study used survey data collected in 2015 because, according to the surveys by the China Internet Network Information Center in 2015 and 2020, penetration rates of major new media applications related to this study (search engines, news applications, and mobile communication apps, including WeChat and microblogs) did not change much between 2015 and 2020. The survey targeted residents in Beijing because Beijing had China's highest internet penetration rate in 2015, at 76.5% [44], which partially excludes the problem of the physical digital divide. Given its vibrant internet activities and high economic level (including gross domestic product per capita), which is comparable to developed countries, the choice of Beijing residents as the sample renders the results useful for scholars in Western countries such as the United States, where most health communication studies take place.

The researchers hired Sojump, a Chinese online panel company, to collect the data because this study only focused on internet users, which allowed the researchers to remove those without internet access and thus the problem of the digital divide. This study used quota sampling to draw participants, which is common for online panel surveys [45]. The quota sampling scheme used age group and gender as the criteria in designing subgroups (an equal number of respondents in the 8 subgroups). Age in this study was divided into the categories of 18 to 25 years, 26 to 30 years, 31 to 40 years, and 41 years and older. The researchers oversampled people older than 30 years because that age is positively related to perception of health risks, which may elicit health information seeking. In the panel, 170,000 participants were from Beijing. For this study, Sojump randomly sent survey invitations in September 2015 to 9500 people within the panel who were registered and confirmed as from Beijing. The data collection process lasted for 10 days and resulted in a sample size of 542. The average age of the respondents was 33 years old, female respondents accounted for 52.0% (282/542) of the sample, and the average monthly income ranged from US $900 to $1200.

Dependent Variables

The outcome variable was the extent to which an individual uses a particular online channel to obtain health information, and this was measured on a 7-point scale. According to preliminary in-depth interviews with Chinese health communication practitioners and users, the researchers pinpointed 9 types of commonly used online health information channels. Given that one channel can be suitable for more than one type of information acquisition behavior, the researchers measured the frequencies of 13 types of online health information channel use (Table 1).

Table 1. Descriptive statistics of online channel use in Beijing, China (N=542).

Category and online health information channels	Mean[a] (SD)
Browsing channels[b]	
WeChat[c] official accounts (ie, health-related official accounts, similar to Facebook pages)	4.58 (1.92)
Microblogs (ie, health microblogs, similar to Twitter)	4.22 (2.05)
Web portals (ie, health section, similar to Yahoo)	3.85 (1.91)
Online professional health sites (similar to WebMD)	3.72 (1.98)
Mobile phone health apps	3.67 (2.05)
Online forums (ie, forums related to a particular disease or health problem)	3.61 (2.02)
Searching channels[b]	
Search engines	5.29 (1.64)
Online encyclopedia sites	4.61 (1.84)
Question-and-answer sites	4.51 (1.85)
Scanning channels[d]	
WeChat Moments (similar to Facebook News Feed)	5.12 (1.66)
Web portals (ie, nonhealth sections)	4.52 (1.67)
Microblogs (ie, nonhealth microblogs)	4.38 (1.81)
Online forums (ie, nonhealth forums)	4.06 (1.82)

[a]Scale of 1 to 7: 1=never, 7=very often.

[b]Browsing channels and searching channels are the channels for active seeking.

[c]WeChat is China's largest mobile messenger service and its functionality is similar to Facebook.

[d]Scanning channels are the channels for incidental exposure, passive exposure, and routine seeking.

When wording the question items, the researchers took into account whether a channel was for active seeking or incidental exposure (scanning). For active-seeking channels, the survey participant was asked, "Over the past year, to what extent have you used the following channel to acquire information or knowledge related to generic health, disease prevention/treatment and healthy living?" [9]. For scanning channels, the survey participant was asked "Over the past year, to what extent did you accidently pay attention to information or knowledge related to generic health, disease prevention/treatment and healthy living while engaged in media tasks other than health information search?" [9]. As shown in Table 1, search engines (active) and WeChat Moments (incidental exposure, similar to Facebook News Feed) were the 2 most frequently used channels (search engines: mean 5.29, SD 1.64; WeChat Moments: mean 5.12, SD 1.66). Online health forums and mobile health apps were the 2 least used channels (health forums: mean 3.61, SD 2.02; health apps: mean 3.67, SD 2.05).

Predictor Latent Variables

This study measured 7 RISP predictors. Table 2 lists specific survey question items and their descriptive statistics.

Table 2. Descriptive statistics of RISP predictor variables (N=542).

Variables and questions	Mean (SD)
Information insufficiency[a] (1-100 scale)	
Perceived current health knowledge: Estimate your knowledge of health, with 1=knowing nothing and 100=knowing everything you could possibly know about health maintenance.	59.88 (17.13)
Sufficiency threshold: This time, using that same scale, estimate how much knowledge you think you need on health maintenance.	73.27 (20.56)
Information subjective norms (1-5 scale)[b]	
Injunctive: My family expects me to seek health knowledge.	3.78 (0.88)
Injunctive: My friends expect me to seek health knowledge.	3.62 (0.94)
Descriptive: People in my life whose opinions I value seek health knowledge.	3.53 (0.86)
Perceived risk (1-5 scale)[c]	
My health may face problems in the next year.	2.51 (1.09)
In the next year, I may possibly suffer from diseases that may impact my job or life.	2.44 (1.18)
In the next year, I am confident about my health (reverse coding).	2.20 (0.97)
Negative affect (1-5 scale)[d]	
How much of the following do you feel about your health? Not worried…Very worried	3.11 (1.05)
How much of the following do you feel about your health? Not anxious…Very anxious	2.71 (1.14)
Relevant channel beliefs (1-5 scale)[e]	
To what extent do you trust health information on web portals?	3.53 (0.74)
To what extent do you trust health information on social media?	3.46 (0.82)
To what extent do you trust health information on mobile phone apps?	3.45 (0.92)
Perceived information-gathering capacities (1-5 scale)[f]	
It is difficult to find health knowledge (reverse coding).	3.39 (1.06)
I don't know where to find health knowledge (reverse coding).	3.37 (1.07)
I have a hard time understanding health knowledge (reverse coding).	3.38 (1.23)
Age	33.01 (9.19)
Gender (female), %	52.0 (—)
Income	3.76 (1.51)

[a]In the questionnaire, the item "current knowledge" was presented first, followed by the item "sufficiency threshold."

[b]α=.70.

[c]α=.81.

[d]α=.79.

[e]α=.61.

[f]α=.82.

Information insufficiency was assessed using 2 items: perceived current health knowledge and information sufficiency threshold [12]. In the analysis, information insufficiency's influence on health information seeking was evaluated by modeling the information sufficiency threshold when controlling for the self-assessment of current health knowledge.

Informational subjective norms included injunctive and descriptive norms [35]. The former refers to important others'

attitudes toward one's behaviors, while the latter refers to important others' behaviors. A total of 3 items were used.

Perceived risks evaluated one's self-assessed probability of suffering from diseases or encountering health problems, which is consistent with previous RISP research [36]. This construct consisted of 3 items with a 5-point scale.

Negative affect assessed one's negative feeling related to their health status [36]. It consisted of 2 items.

Relevant channel beliefs evaluated one's belief that online channels are capable of supplying trustworthy information [35]. In this study, the 3 items specifically tapped into the dimension of trust, since Chinese online media are often criticized for containing misinformation related to health.

Perceived information-gathering capacities assessed one's belief that they are capable of accessing and understanding health information [35]. Three 5-point items were used to measure this concept.

Results

Overview

The researchers analyzed the data using structural equation modeling techniques. Initially, the researchers reduced the use of 13 online channels into 3 groups of variables, which were used as endogenous variables [32]. Then, the structural model was built. The lavaan package in R was used to test hypotheses.

Measurement Model

Three measurement models were constructed and compared (Multimedia Appendix 1 illustrates the specific measurement model–building procedures). The third measurement model had the best model fit (Table 3). The third measurement model classified the 13 channel use variables into 3 groups: browsing, searching ,and scanning channels. Additionally, the third measurement model correlated the error terms of 6 pairs of variables of channel use that have moderate to high levels of correlation due to shared technical platforms (Multimedia Appendix 2). For instance, people who prefer health sections of web portals are likely to have the habit of using web portals, which raises their chances of encountering health messages when using web portals for tasks unrelated to health.

Table 3. Model fit statistics.[a]

Model	Chi-square (*df*)	χ^2/df	RMSEA[b]	CFI[c]	SRMR[d]
Measurement model					
Model 1: Baseline (2-factor)[e]	1149.1 (303)	3.79	0.072	0.86	0.059
Model 2. Revised (3-factor)[f]	904.8 (296)	3.06	0.062	0.90	0.052
Model 3: Revised + correlated error	638.0 (290)	2.20	0.047	0.94	0.049
Structural model					
Model 1: Baseline conceptual	923.0 (343)	2.69	0.056	0.91	0.081
Model 2: Revised	931.7 (351)	2.65	0.055	0.91	0.083

[a]Recommended cutoff points for model fit indices [46-48]: SRMR of <0.08; RMSEA of <0.08; CFI of >0.90 (ideally CFI of ≥0.95); χ^2/df of <3. Hu and Bentler [47] suggest a 2-index presentation strategy, recommending a RMSEA of 0.06 or lower and a SRMR of 0.09 or lower.

[b]RMSEA: root mean square error of approximation.

[c]CFI: comparative fit index.

[d]SRMR: standardized root mean residual.

[e]2-factor: active seeking and scanning channels.

[f]3-factor: searching, browsing, and scanning channels.

The results from the measurement model suggested that the use of online health information channels can be divided into 3 categories: browsing, searching, and scanning channels (Table 1). Browsing channels consist of health sections of web portals (B=1.00; β=0.72; *P*<.001), professional health sites (B=1.01; β=0.70; *P*<.001), microblogs (B=1.01; β=0.67; *P*<.001), WeChat official accounts (B=0.91; β=0.64; *P*<.001), mobile health apps (B=1.08; β=0.73; *P*<.001), and online health forums (B=1.09; β=0.75; *P*<.001).

Searching channels include search engines (B=1.00; β=0.65; *P*<.001), online encyclopedia sites (B=1.35; β=0.79; *P*<.001), and question-and-answer sites (B=1.35; β=0.78; *P*<.001).

Scanning channels encompass incidental exposure to health information on web portals (B=1.00; β=0.73; *P*<.001), microblogs (B=1.02; β=0.69; *P*<.001), WeChat Moments (B=0.68; β=0.50; *P*<.001), and online forums (B=1.14; β=0.76; *P*<.001).

Structural Model

The bottom of Table 3 presents the model fit indices for the structural model. Two structural models, conceptual and revised (removing the nonsignificant paths), were compared. The popular model fit indices were roughly the same. The revised model (Figure 2) was retained, since the likelihood test suggested that the 2 competing models were not significantly

different ($\Delta\chi^2_1$=8.7, P=.37). Figure 2 presents test results of H1 to H10.

H1 examined the extent to which the information sufficiency threshold was associated with online channel choice. According to the results, the information sufficiency threshold was positively associated only with the use of the online searching channel (B=0.09; β=0.01; *P*<.001) and was negatively related to the use of online browsing channels (B=–0.004; β=–0.064; P=.03). H1a and H1b were supported. However, the information sufficiency threshold was not related to the use of scanning channels, and H1c was not supported. H2 analyzed in what way current knowledge was related to online channel selection. It was found that current knowledge was only positively associated with the use of online browsing channels (B=0.007; β=0.09; P=.003), not searching (H2a) and scanning (H2c) channels. Thus, only H2b was supported.

H3 examined the positive association between current knowledge level and perceived information-gathering capacities and was supported (B=0.01; β=0.19; *P*<.001). H4 examined the relationship between perceived information-gathering capacities and online channel selection. It was found that perceived information-gathering capacities had negative associations with online scanning channel use (B=–0.39; β=–0.26; *P*<.001). Thus, H4b was supported. Conversely, perceived information was not related to the use of online searching and browsing channels, and H4a and H4b were not supported.

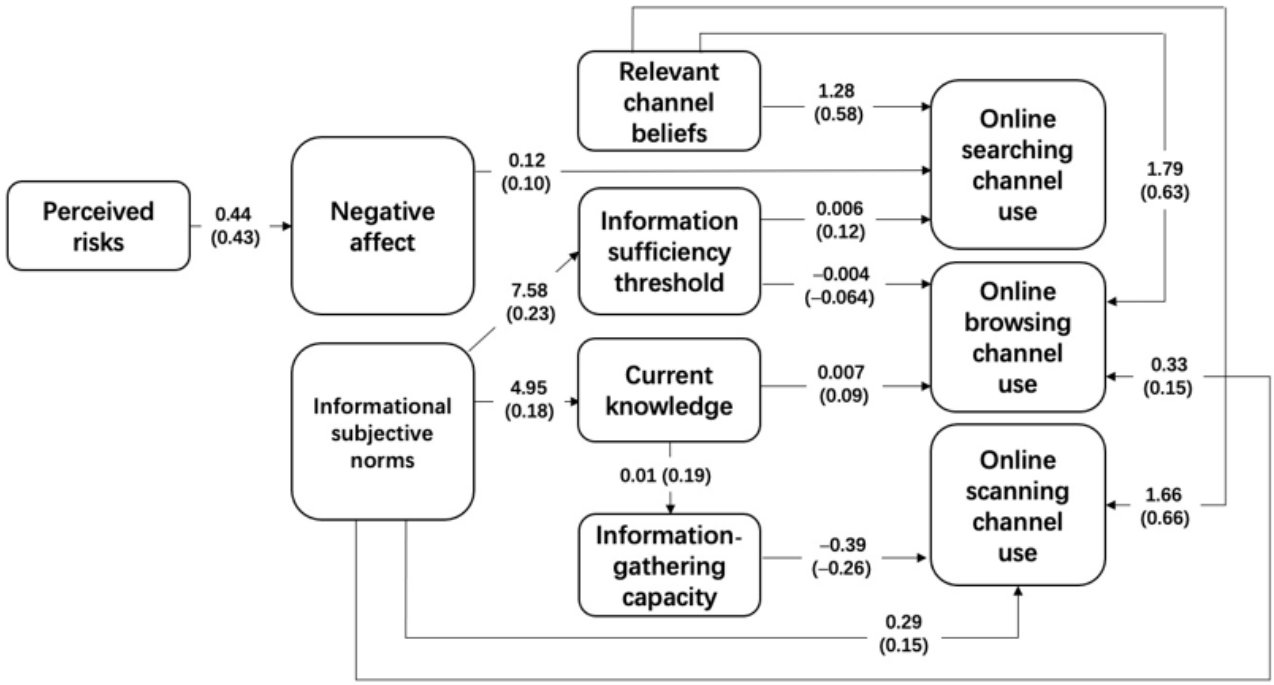
H5 predicted the positive association of perceived risks with negative affect, and it was supported (B=0.44; β=0.43; *P*<.001).

H6 predicted that negative affect would be positively associated with the information sufficiency threshold. However, this was not supported. H7 examined the extent to which negative affect was associated with channel choice. The results suggested that negative affect was positively associated only with online searching channel use (B=0.11; β=0.10; *P*=.02) and not the use of the other 2 types of online channels. Thus, only H7a was supported.

H8a and H8b tested the positive relationship between informational subjective norms and the information sufficiency threshold and current knowledge, both of which were supported. The results suggested that informational subjective norms were positively associated with the information sufficiency threshold (B=7.58; β=0.23; *P*<.001) (H8a) and current knowledge (B=4.95; β=0.18; *P*<.001) (H8b). H9 examined the relationship between informational subjective norms and online channel selection. The results revealed that informational subjective norms were directly related to online browsing (B=0.33; β=0.15; *P*=.004) (H9b) and scanning channels (B=0.29; β=0.15; *P*=.007) (H9c). However, informational subjective norms did not have a direct association with the use of searching channels (H9a).

H10 tested whether relevant channel beliefs were related to online channel use, which was supported. Additionally, the sizes of the coefficients were roughly the same (for searching channels [H10a]: B=1.28; β=0.58; *P*<.001; for browsing channels [H10b]: B=1.79; β=0.63; *P*<.001; for scanning channels [H10c]: B=1.66; β=0.66; *P*<.001).

Figure 2. Revised model. The numbers in parentheses are standardized coefficients and the numbers preceding the parentheses are unstandardized coefficients.



Discussion

Overview

This study described how people in Beijing, China, use online channels to acquire health information. The results suggested that the 13 commonly used online health information channels can be divided into 3 categories: browsing, searching, and scanning channels (Table 1). These were also corroborated by the underlying motivational mechanism.

Diverse Use of Online Health Information Channels

This study found that Chinese people use online health information channels to different extents. Along with search engines, China's largest messenger service, WeChat (WeChat Moments, a type of scanning channel), was reported as one of the 2 most frequently used channels. Other frequently used channels included WeChat official accounts (similar to Facebook pages), online encyclopedia sites, question-and-answer sites, microblogs (browsing and scanning), and web portals (scanning). The descriptive statistics clearly show that although search engines were the most frequently used, the importance of social media in online health information acquisition in China outweighs that of the channels run by institutions, which is in line with previous studies that found that people in East Asia prefer social media when seeking health information [17].

Patterns of Online Health Information Channel Use

The confirmatory factor analysis further revealed that the use of online health information channels can be classified into 3 categories. From the perspective of the theory of channel complementarity [8], the variable clustering pattern may hint that people use different online channels with similar functions complementarily to obtain health information. Search engines, online encyclopedias, and question-and-answer sites are grouped together. When people use search engines, they use specific keywords to actively search for information related to specific health concerns. Online encyclopedia sites are user-generated reference "books" in which people also commonly use specific keywords to query information. Additionally, question-and-answer site users with specific questions in mind actively solicit answers from fellow users, which is similar in nature to user behaviors on search engines and online encyclopedias. Therefore, search engines, online encyclopedias, and question-and-answer sites are online searching channels that users mainly use to search for answers related to health concerns. An alternative explanation might be that, from researchers' anecdotal observations, search engines in China often return top results linking users to online encyclopedia sites and question-and-answer sites, which suggests that these two types of sites successfully apply search engine optimization strategies.

Web portals (eg, health sections), official accounts on WeChat, microblogs (eg, health microblog accounts), professional health sites, mobile health apps, and online health forums are combined into a separate group: online browsing channels. The content and layout of the content on the first 5 online channels are commonly prepared by professional editors; users normally follow the structure of information prepared by the publisher [10]. It is also noted that online health forums join the other 5 channels mentioned earlier. While online health forums allow users to solicit answers about specific questions, many users are usually spectators scrolling through the posts and threads of others. Thus, the use of online health forums is more similar to the use of health sections of web portals than the use of search engines and question-and-answer sites.

Updates on WeChat Moments and microblogs, web portals (eg, general news section), and online forums not related to health are grouped together as online scanning channels. Users encounter health information on these channels when engaged in tasks other than active health information seeking.

Differences in Sociopsychological Mechanisms for Channel Choice

This study further pinpointed the variation in motivation to use different types of health information channels from the perspective of RISP [12,33] and showed the validity of RISP in explaining information channel selection. The differences in sociopsychological mechanisms for channel selection corroborated the validity of the results generated by confirmatory factor analysis.

The use of searching channels, including search engines, question-and-answer sites, and online encyclopedia sites, is motivated by the intention to acquire more health knowledge (information sufficiency threshold) and by negative affect. Additionally, online searching channel use is indirectly caused by perceived risks via negative affect and indirectly influenced by informational subjective norms via the information sufficiency threshold. An individual who chooses search engines to search for specific disease-related keywords to reduce uncertainties is likely to feel threatened by health problems and have a stronger need for health knowledge.

The use of online scanning channels, including WeChat Moments, microblogs, online forums, and web portals, is driven primarily by informational subjective norms. An individual who pays attention to health information or knowledge on online scanning channels is unlikely to be concerned about their health status or increasing their knowledge. The use of online scanning channels to obtain health knowledge is attributed more to pressure from the individual's interpersonal social network.

Similar to the use of online scanning channels, the use of online browsing channels is also directly driven by influence from one's social network. Additionally, informational subjective norms predict the use of browsing channels via current knowledge. Browsing content on online professional health sites, for instance, does not necessarily mean that an individual can immediately benefit from the content read. Only if the individual understands the potential long-term benefits of that information or knowledge can they take the time to process the information. A higher level of current knowledge may help individuals be aware of the long-term benefits of health knowledge accruement. In other words, the use of browsing channels is comparable to reading lengthy books, while the use of searching channels is similar to looking up the definition of a term in a dictionary.

Interestingly, information-gathering capacity was negatively associated with the extent to which respondents used scanning channels. Being able to locate health knowledge might prevent users from using scanning channels because they might believe in other, more efficient routes of obtaining health information, such as those under the category of searching or browsing channels.

Practical Implications

The results of the study suggest that health communication practitioners and scholars should measure "internet," "new media," and "online media" more precisely instead of simply asking the public about the frequency of internet use in health information acquisition. The measurement of generic internet use may cause too much error, and it lends no support to media planning in a public health promotion campaign.

More importantly, as contemporary health care consumers reside in a multichannel environment [49], health communication practitioners and scholars may consider developing more appropriate methods of classifying these channels to better manage them. As mentioned above, this study devises a new classification scheme based on health information–seeking behaviors. Practitioners may consider categorizing online channels using the new scheme generated by this study, which may better cater to the needs of individuals planning consumer-centered health communication campaigns.

The channel choice pattern and underlying sociopsychological mechanisms generate useful insights to improve health information strategies as well. We found that the use of online searching channels was uniquely driven by the need for more health knowledge and by stronger negative affect. In other words, people who use search engines, online encyclopedias, and question-and-answer sites may readily accept the answers found through these channels and use the knowledge to guide their health behaviors if they find the answers plausible. However, these 3 types of channels are often dominated by nonmedical professionals and possibly contain misinformation [50], so professional health agencies may consider establishing closer partnerships with these channels.

It was found that the use of online browsing channels was uniquely driven by self-reported current knowledge level, which implies that people with higher health literacy are more likely to choose browsing channels. Since most browsing channels are run by institutions, these channels and websites should focus on the provision of more advanced health knowledge. The basics should be left to sites such as online encyclopedias.

This study also found that the likelihood of using online scanning channels was primarily influenced by informational subjective norms. Additionally, health culture, which makes health a shared value, is beneficial to people's health behaviors [51]. This means that if we could build a culture of health conducive to the habitual acquisition of health knowledge, users would be likely to process at least some health information encountered on online scanning channels. However, it should be noted that scanning channels may also contain a considerable amount of health content generated by nonprofessionals. Accuracy of online health information is always a concern [50], so practitioners should involve themselves as much as possible in such channels.

Limitations and Future Studies

This study is not without limitations. Constrained by budgets, this study used an online panel to collect data. Although the sample covered a range of age groups, this study did not sample enough older adults, which to some extent limited the generalization of the results to the entire population. Additionally, online panels consist of respondents who are paid to respond to the survey questions. These respondents may differ from the general population, which may bias the results of the study.

Since this study revealed that motivations behind using different types of channels differ, future studies may further explore the differences in the impacts of online channel use on health knowledge gains and behavioral changes. Moreover, future research is advised to further explore how people process information when using different types of online health information channels. The fourfold typology of information seeking and processing suggests that different information-seeking strategies may possibly entail different styles of information processing. Researchers do not examine online health information channel use and selection for the sake of merely understanding channel selection; instead, the ultimate goal of such research is to have a better understanding of how different channel choices influence changes in people's cognition, attitudes, and behaviors.

Acknowledgments

This research was supported by National Social Science Foundation of China (No. 20&ZD319).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Model building.
[DOCX File , 18 KB - jmir_v23i3e24945_app1.docx]

Multimedia Appendix 2

Correlation tables.
[DOCX File , 22 KB - jmir_v23i3e24945_app2.docx]

References

1. Fox S, Purcell K. Chronic disease and the internet. Pew Research Center. 2010 Apr 24. URL: http://www.pewinternet.org/2010/03/24/chronic-disease-and-the-internet/ [accessed 2021-03-03]
2. Cao W, Zhang X, Xu K, Wang Y. Modeling online health information-seeking behavior in China: The roles of source characteristics, reward assessment, and internet self-efficacy. Health Commun 2016 Sep;31(9):1105-1114. [doi: 10.1080/10410236.2015.1045236] [Medline: 26861963]
3. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
4. Dunwoody S, Griffin R. The role of channel beliefs in risk information seeking. In: Arvai J, Rivers III L, editors. Effective Risk Communication. London, United Kingdom: Taylor & Francis; 2014:220-233.
5. Flanagin A. Internet use in the contemporary media environment. Hum Commun Res 2001 Jan 01;27(1):153-181. [doi: 10.1093/hcr/27.1.153]
6. Cormode G, Krishnamurthy B. Key differences between Web 1.0 and Web 2.0. First Monday 2008 Apr 25;13(6):2125. [doi: 10.5210/fm.v13i6.2125]
7. Keselman A, Logan R, Smith CA, Leroy G, Zeng-Treitler Q. Developing informatics tools and strategies for consumer-centered health communication. J Am Med Inform Assoc 2008 Jul;15(4):473-483 [FREE Full text] [doi: 10.1197/jamia.M2744] [Medline: 18436895]
8. Dutta-Bergman M. Media use theory and internet use for health care. In: Murero M, Rice RE, editors. The Internet and Health Care: Theory, Research, and Practice. Mahwah, NJ: Lawrence Erlbaum Associates; 2006:83-106.
9. Niederdeppe J, Hornik RC, Kelly BJ, Frosch DL, Romantan A, Stevens RS, et al. Examining the dimensions of cancer-related information seeking and scanning behavior. Health Commun 2007;22(2):153-167. [doi: 10.1080/10410230701454189] [Medline: 17668995]
10. Zhang Y, Wang P, Heaton A, Winkler H. Health information searching behavior in MedlinePlus and the impact of tasks. 2012 Presented at: 2nd ACM SIGHIT International Health Informatics Symposium; Jan 28-30, 2012; Miami, Florida. [doi: 10.1145/2110363.2110434]
11. Marchionini G, Komlodi A. Design of interfaces for information seeking. Ann Rev Inf Sci Technol 1998;33:89-130.
12. Griffin RJ, Dunwoody S, Neuwirth K. Proposed model of the relationship of risk information seeking and processing to the development of preventive behaviors. Environ Res 1999 Feb;80(2 Pt 2):S230-S245. [doi: 10.1006/enrs.1998.3940] [Medline: 10092438]
13. Brashers DE, Goldsmith DJ, Hsieh E. Information seeking and avoiding in health contexts. Hum Commun Res 2002 Apr;28(2):258-271. [doi: 10.1111/j.1468-2958.2002.tb00807.x]
14. Fox S, Duggan M. Health online 2013. Pew Research Center. 2013 Jan 15. URL: https://www.pewresearch.org/internet/2013/01/15/health-online-2013/ [accessed 2021-03-03]
15. Yu G, He R. Big data application in health communication: Content, consequences and challenges. Chin J Editors 2013;6:20-22.
16. Dai F, Liu Y, Su Y, Yi X. Research on patients' acquisition of health information and utilization of medical services under the network environment. J Med Postgrad 2014;27(5):517-520.
17. Song H, Omori K, Kim J, Tenzek KE, Morey Hawkins J, Lin W, et al. Trusting social media as a source of health information: online surveys comparing the United States, Korea, and Hong Kong. J Med Internet Res 2016 Mar 14;18(3):e25 [FREE Full text] [doi: 10.2196/jmir.4193] [Medline: 26976273]
18. Case D. Looking for information: A survey of research on information seeking, needs, and behavior. London, United Kingdom: Academic Press; 2007.
19. Bates MJ. Toward an integrated model of information seeking and searching. N Rev Inf Behav Res 2002;3(1):1-15.
20. Zhou X, Cai W. College students online health information seeking behavior patterns and influencing factors. Inf Documentation Serv 2014;21(4):50-55.
21. Ruppel EK. Scanning health information sources: applying and extending the comprehensive model of information seeking. J Health Commun 2016;21(2):208-216. [doi: 10.1080/10810730.2015.1058438] [Medline: 26716985]
22. Johnson JD, Meishcke H. Differences in evaluations of communication channels for cancer-related information. J Behav Med 1992 Oct;15(5):429-445. [doi: 10.1007/bf00844940]
23. Johnson JD, Meischke H. Cancer information: women's source and content preferences. J Health Care Mark 1991 Mar;11(1):37-44. [Medline: 10110080]
24. Katz E, Blumler JG, Gurevitch M. Utilization of mass communication by the individual. In: Katz E, Blumler JG, Gurevitch M, editors. The uses of mass communications: Current perspectives on gratifications research. Beverly Hills, CA: Sage; 1974:19-32.
25. Ruggiero TE. Uses and gratifications theory in the 21st century. Mass Commun Soc 2000 Feb;3(1):3-37. [doi: 10.1207/s15327825mcs0301_02]

26. Ajzen I. Attitudes, personality, and behavior. Chicago, IL: Dorsey; 1988.
27. Chaiken S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. J Pers Soc Psychol 1980 Nov;39(5):752-766. [doi: 10.1037/0022-3514.39.5.752]
28. Johnson EJ, Tversky A. Affect, generalization, and the perception of risk. J Pers Soc Psychol 1983 Jul;45(1):20-31. [doi: 10.1037/0022-3514.45.1.20]
29. Kosicki G, McLeod JM. Learning from political news: Effects of media images and information processing strategies. In: Kraus S, editor. Mass communication and political information processing. Hillsdale, NJ: Erlbaum; 1990:69-83.
30. Yang JZ, Huang J. Seeking for your own sake: Chinese citizens' motivations for information seeking about air pollution. Environ Commun 2018 Jan 05;13(5):603-616. [doi: 10.1080/17524032.2017.1397041]
31. Yang ZJ. Too scared or too capable? Why do college students stay away from the H1N1 vaccine? Risk Anal 2012 Oct;32(10):1703-1716. [doi: 10.1111/j.1539-6924.2012.01799.x] [Medline: 22404650]
32. Yang ZJ, McComas KA, Gay G, Leonard JP, Dannenberg AJ, Dillon H. Information seeking related to clinical trial enrollment. Commun Res 2010 Sep 08;38(6):856-882. [doi: 10.1177/0093650210380411]
33. Griffin R, Dunwoody S, Yang Z. Linking risk messages to information seeking and processing. In: Salmon CT, editor. Communication Yearbook 36. New York, NY: Taylor & Francis; 2012:332-362.
34. Rose KM, Toman E, Olsen CS. Public use of information about smoke emissions: application of the risk information seeking and processing (RISP) model. Can J Forest Res 2017 Nov;47(11):1527-1537. [doi: 10.1139/cjfr-2017-0099]
35. Yang ZJ, Kahlor L, Li H. A United States-China comparison of risk information–seeking intentions. Commun Res 2013 Mar 12;41(7):935-960. [doi: 10.1177/0093650213479795]
36. Yang ZJ, Aloe AM, Feeley TH. Risk information seeking and processing model: a meta-analysis. J Commun 2014 Jan 07;64(1):20-41. [doi: 10.1111/jcom.12071]
37. Radecki CM, Jaccard J. Perceptions of knowledge, actual knowledge, and information search behavior. J Exp Soc Psychol 1995 Mar;31(2):107-138. [doi: 10.1006/jesp.1995.1006]
38. Schmidt JB, Spreng RA. A proposed model of external consumer information search. J Acad Marketing Sci 1996 Jun 01;24(3):246-256. [doi: 10.1177/0092070396243005]
39. Mishra SK, Kumar M. How mutual fund investors' objective and subjective knowledge impacts their information search and processing behaviour. J Financ Serv Mark 2011 May 26;16(1):27-41. [doi: 10.1057/fsm.2011.1]
40. Griffin RJ, Zheng Y, ter Huurne E, Boerner F, Ortiz S, Dunwoody S. After the flood. Sci Commun 2008 Mar;29(3):285-315. [doi: 10.1177/1075547007312309]
41. Yang ZJ, Kahlor L. What, me worry? The role of affect in information seeking and avoidance. Sci Commun 2012 May 24;35(2):189-212. [doi: 10.1177/1075547012441873]
42. Fung TKF, Griffin RJ, Dunwoody S. Testing links among uncertainty, affect, and attitude toward a health behavior. Sci Commun 2018 Jan 12;40(1):33-62. [doi: 10.1177/1075547017748947]
43. Kim HK, Kim Y. Risk information seeking and processing about particulate air pollution in South Korea: the roles of cultural worldview. Risk Anal 2018 Nov 08;39(5):1071-1087. [doi: 10.1111/risa.13231]
44. The 36th statistical report on internet development in China. China Internet Network Information Center. 2017. URL: http://cnnic.com.cn/IDR/ReportDownloads/201601/P020160106496544403584.pdf [accessed 2021-03-03]
45. Callegaro M, Baker R, Bethlehem J, Göritz A, Krosnick J, Lavrakas P. Online panel research: A data quality perspective. Hoboken, NJ: Wiley; 2014.
46. Byrne BM. Structural equation modeling with AMOS: Basic concepts, applications, and programming. Mahwah, NJ: Lawrence Erlbaum; 2001.
47. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equation Modeling Multidisciplinary J 1999;6(1):1-55. [doi: 10.1080/10705519909540118]
48. Kline RB. Principles and practice of structural equation modeling. New York, NY: Guilford; 1998.
49. Krivich M. The healthcare consumer lives in a multichannel environment; the response is? LinkedIn. URL: https://www.linkedin.com/pulse/healthcare-consumer-lives-multichannel-environment-michael-j/ [accessed 2021-03-03]
50. Kim SU, Syn SY. Credibility and usefulness of health information on Facebook: A survey study with US college students. Inf Res 2016;21(4):1-27.
51. Chandra A, Acosta J, Carman KG, Dubowitz T, Leviton L, Martin LT, et al. Building a national culture of health: background, action framework, measures, and next steps. Rand Health Q 2017 Jan;6(2):3 [FREE Full text] [Medline: 28845341]

Abbreviations

RISP: risk information seeking and processing

Edited by G Eysenbach; submitted 11.10.20; peer-reviewed by C Li, W Pan, J Lander; comments to author 02.11.20; revised version received 31.12.20; accepted 27.02.21; published 09.03.21.

Please cite as:
Zhang D, Shi Z, Hu H, Han G(
Classification of the Use of Online Health Information Channels and Variation in Motivations for Channel Selection: Cross-sectional Survey
J Med Internet Res 2021;23(3):e24945
URL: https://www.jmir.org/2021/3/e24945
doi:10.2196/24945
PMID:33687342

Original Paper

Share to Seek: The Effects of Disease Complexity on Health Information–Seeking Behavior

Ashwag Alasmari[1,2], PhD; Lina Zhou[3], PhD

[1]Computer Science Department, King Khalid University, Abha, Saudi Arabia
[2]University of Maryland, Baltimore County, Baltimore, MD, United States
[3]University of North Carolina at Charlotte, Charlotte, NC, United States

Corresponding Author:
Ashwag Alasmari, PhD
University of Maryland, Baltimore County
1000 Hilltop Cir
Baltimore, MD, 21250
United States
Phone: 1 (410) 455 1000
Email: ashwag1@umbc.edu

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2022/9/e42949

Abstract

Background: Web-based question and answer (Q&A) sites have emerged as an alternative source for serving individuals' health information needs. Although a number of studies have analyzed user-generated content in web-based Q&A sites, there is insufficient understanding of the effect of disease complexity on information-seeking needs and the types of information shared, and little research has been devoted to the questions concerning multimorbidity.

Objective: This study aims to investigate seeking of health information in Q&A sites at different levels of disease complexity. Specifically, this study investigates the effects of disease complexity on information-seeking needs, types of information shared, and stages of disease development.

Methods: First, we selected a random sample of 400 questions separately from each of the Q&A sites: Yahoo Answers and WebMD Answers. The data cleaning resulted in a final set of 624 questions from the two sites. We used a mixed methods approach, including qualitative content analysis and quantitative statistical analysis.

Results: The one-way results of ANOVA showed significant effects of disease complexity (single vs multimorbid disease questions) on two information-seeking needs: diagnosis ($F_{1,622}$=5.08; P=.02) and treatment ($F_{1,622}$=4.82; P=.02). There were also significant differences between the two levels of disease complexity in two stages of disease development: the general health stage ($F_{1,622}$=48.02; P<.001) and the chronic stage ($F_{1,622}$=54.01; P<.001). In addition, our results showed significant effects of disease complexity across all types of shared information: demographic information ($F_{1,622}$=32.24; P<.001), medical diagnosis ($F_{1,622}$=11.04; P<.001), and treatment and prevention ($F_{1,622}$=14.55; P<.001).

Conclusions: Our findings present implications for the design of web-based Q&A sites to better support health information seeking. Future studies should be conducted to validate the generality of these findings and apply them to improve the effectiveness of health information in Q&A sites.

(J Med Internet Res 2021;23(3):e21642) doi:10.2196/21642

KEYWORDS

health information consumers; multimorbidity; information searching; information seeking; disease development

Introduction

Background

Reports from the Pew Research Center indicate that an increasing number of people use web-based services to obtain health information, rising from 25% of Americans in 2000 to 72% in 2014 [1]. Users look for specific diseases and treatments or other people with similar health problems, usually to diagnose themselves or others. These web-based resources range from general search engines to specific sites devoted to health information. One potential resource that could meet the health information needs of many internet users is question and answer (Q&A) sites.

Web-based Q&A sites allow health information consumers (HICs) to post questions for other users to answer [2]. Q&A sites may focus on specific topics or be more general in nature. Another characteristic distinction between different types of Q&A sites is whether the posted answers are curated by experts (expert curated) or by other users (community based). In an expert-curated site, experts are considered an essential part of the community and their answers are ranked first, followed by answers posted by other contributors. In a community-based site, HICs seek support from peers with similar conditions, and the answers are featured typically based on the total number of community votes.

The motives of users participating in Q&A sites, especially in the long term, have been explained based on social theories [3]. According to the social exchange theory, individuals exchange valuable resources through contacts to receive something that benefits both parties [4]. This reciprocity encourages social interactions in a co-operative fashion between individuals, even when the rewards are not tangible. Individuals may also share information with someone believing that a third party could offer similar help in the future when needed, and this is termed generalized reciprocity [3]. Beyond finding solutions to their problems, those who seek information possess other benefits such as meta-knowledge (becoming aware of further resources such as people, databases, or documents), problem reformulation, acknowledgment and legitimation from respected people, and validation of their plans or solutions [5]. HICs who participate in web-based Q&A communities want to better understand their condition and receive treatment options from professionals [6], evaluate their health with regard to risks and prevention of further disease [6], or communicate with patients who share similar experiences [7].

Several studies have examined the information-seeking behaviors of HICs based on the characteristics of their questions and information needs [8,9]. A study examining cancer-related topics on Yahoo Answers found that HICs provided rich information about their problems, emotions, social relationships, and life situations in their questions [8]. The findings not only highlighted the complexity of health-related issues but also indicated that there are multiple reasons why people would search on the web for answers. In HICs with cancer, for example, those who post during their cancer diagnosis or treatment ask mainly for advice. Survivors seemed to share personal narratives, and terminal patients sought acknowledgment and validation of their choices [8]. In addition, previous studies have demonstrated that the language used by HICs is based not only on controlled vocabularies developed for lay consumers, such as consumer health vocabularies, but also on the terminology used by health professionals, such as SNOMED Clinical Terms [9,10]. Another stream of research focused on understanding web-based health questions by extracting the linguistic features of the text. For example, a study examined health questions on the topic of eating disorders in Yahoo! Answers for linguistic style and sentiment [11].

One major limitation of previous studies is that they focused on single-disease entities in isolation, such as cancer [8,12,13] or diabetes [9]. These are complex diseases that require careful management, especially when they are associated with the presence of other diseases. The population of patients with multiple chronic conditions, also known as multimorbidity, has increased immensely over the past few decades [14]. Multimorbidity negatively impacts a patient's quality of life, hospitalization, and mortality [15]. Patients with multimorbidity face unique challenges because they are expected to distinguish between the symptoms of different diseases and make sense of the abundance of information obtained from different web-based health resources [16,17]. Despite the rapid growth of online health communities, most of these communities cater to the management of single chronic diseases and are not responsive to the needs of HICs with multimorbidity [16,18,19]. The effect of disease complexity in its entire spectrum (single and multiple diseases) on information-seeking needs and types of information shared by HICs appears to be only partially investigated in the literature.

Objectives

This study aims to investigate health information sharing in web-based Q&A platforms with a specific focus on the effect of disease complexity, namely the characteristics of multiple-disease health questions as compared with characteristics of single-disease health questions. There is little understanding of the effect of disease complexity on the characteristics of health questions, and we do not know how deeply these differences run. More specifically, this paper aims to answer the following research questions: Are there any differences between questions relating to a single disease compared with those relating to multimorbidity in web-based Q&A platforms? How deep do these differences run in terms of information-seeking needs, types of information shared, and stages of disease development?

Methods

Disease Topic Selection

We chose kidney disease as the health topic. According to statistics from the National Institute of Diabetes and Digestive and Kidney Diseases, kidney disease ranges from simple conditions, such as small stones, to more serious illnesses such as chronic kidney disease, the prevalence of which is estimated to be 14% in the general population [20]. Chronic kidney disease has significant complications such as high blood pressure, cardiovascular disease, anemia, weak bones, poor nutritional health, and neuropathy [21]. At the same time, other chronic

conditions increase the risk of chronic kidney disease, such as diabetes and high blood pressure. Thus, multimorbidity among patients with kidney disease is extremely high, which would inevitably have a significant impact on their ability to identify, evaluate, and manage their health [22-26].

Q&A Sites

We collected data from an expert-curated Q&A site, WebMD Answers, and a community-based site, Yahoo! Answers. WebMD is one of the most influential web-based health sites [27,28], and users were able to post questions for certified health experts to answer in the Q&A section, which covered more than 900 health topics. We collected all posts from August 2008 until the closure of WebMD Answers in 2018 and its successor, that is, the Questions & Answers A-Z section that offers preset answers and questions.

Yahoo! Answers features health as one of the top-level categories; therefore, it was selected as a community-based Q&A site. Since its creation in December 2005, Yahoo! Answers has become a popular internet reference site worldwide, and it is the most frequented community Q&A site in the United States. As of June 2019, the site ranked ninth in global internet traffic and engagement over the past 90 days and seventh in the United States [29]. We collected data from this site for a period of 9 years (2006-2015).

Sampling Process, Data Collection, and Preparation

Given the selection of kidney disease, we screened questions based on the following key terms: kidney, kidney infection, kidney stone, kidney cancer, kidney disease, chronic kidney disease, dialysis, kidney failure, renal artery stenosis, and renal cell carcinoma. These key terms were selected as they directly refer to kidney conditions or early signs of chronic kidney disease [30]. Using an application programming interface, we sampled 400 random questions from Yahoo! Answers and 400 questions from WebMD related to the abovementioned kidney-related key terms.

We manually removed noise such as advertisements, irrelevant questions (non–kidney-related, nonhuman subjects, or student projects), or posts that did not have an actual question. After cleaning, we had 316 questions from Yahoo Answers and 308 from WebMD Answers. In total, the data cleaning process resulted in a final set of 624 questions. For each of the included questions, we extracted the title, descriptions, date of posting, categories under which the questions were posted, number of answers, and the answers themselves.

The data sets contained both qualitative (the actual content) and quantitative information (such as dates of posting and replies, demographics of participants, topic, and length and number of questions). In this study, we focused on the questions themselves. Therefore, a mixed methods approach was necessary for this analysis.

Content analysis is widely used as a qualitative research method for analyzing questions and answers from web-based Q&A websites [3]. The content of each post is coded into a set of categories, and analysis is performed based on these categories and their frequencies. In this study, we used directed content analysis, where the analysis started with findings from previous research as guidance for the initial codes [31]. The qualitative data were transformed into quantitative data through deductive analysis based on predetermined frameworks and themes and were then analyzed quantitatively. Quantitative data analysis allowed for comparisons of groups and for examining the effect of covariates, in this case, disease complexity.

Dependent Variables and Independent Variables

The independent variable was disease complexity, with two levels: single and multiple (multimorbidity). We measured the variables based on the number of diseases described in the health questions. The health question was labeled as *single* if it was related to a single disease (kidney only) and otherwise labeled as *multiple*. We considered the following dependent variables: information-seeking needs, stages of disease development, and types of shared information.

We contextualized the health questions by adapting relevant findings from a previous study on health information seeking [32]. The three main areas of interest for our analysis included health information–seeking needs, stages of disease development, and information shared (personal and medical). Importantly, in recognition that chronic diseases are more complex, we introduced a comprehensive disease development stage by adapting a framework of chronic illness development [33]. We coded the stages of disease development using the following Zhang and Corbin frameworks, and the coders were asked to decide which disease stage the health questions are related to. The information-seeking needs and types of information shared included several variables coded as binary (present or absent). The following are detailed descriptions of our coding schema.

Information-Seeking Needs

HICs ask health-related questions to address specific information needs [32]:

- Symptom: to gain an understanding of the symptoms of a kidney or any other related disease.
- Diagnosis: to confirm the nature of a certain disease.
- Causes: to figure out the causes of the disease.
- Prognoses: to inquire about the hypothetical effect of a disease.
- Treatment: to explore treatment alternatives to kidney disease.
- Supplements and lifestyle: to explore lifestyle and diet in people with kidney disease and use different supplements.
- Information sources, medical profession, and related types of information: to look for medical experts in the field and any kind of resources to fulfill HIC information needs.
- Drug interaction: to ask for more details about unfavorable and unexpected signs, symptoms, or diseases associated with the use of a drug without any judgment about the causality or relationship to drug use.
- Similar experiences: to connect to patients with similar conditions.

Stage of Disease Development

Diagnosing and treating a disease or condition is an ongoing process. HICs at different stages in this process often have different levels of information needs [13]. HICs may also display different information-seeking behaviors based on the nature and extent of their needs. We adopted a disease development model consisting of 8 stages [32]: (1) being healthy; (2) self-diagnosed as being ill; (3) before having a medical test or checkup; (4) after being diagnosed or self-diagnosed as ill; (5) before treatment (such as surgery or medication); (6) during treatment (including medications or exercise); (7) after treatment; and (8) when the disease becomes chronic or reaches the terminal stage.

Multimorbidity is strongly associated with chronic disease. Accordingly, we also drew on the stages of chronic illness in understanding the questions of HICs who are chronically ill. To this end, stages of disease development were extended to include the chronic illness trajectory framework [33] for clinician use in nursing care and chronic illness management. This trajectory framework was built based on the idea that the course of chronic conditions varies and changes over time. It consists of 9 stages [33]: (1) pretrajectory, before disease onset; (2) trajectory onset, appearance of symptoms and diagnosis; (3) stability, condition and symptoms are under control. Everyday life is unaffected, illness management is home-centered, and hospitalization is not required; (4) unstable, condition and symptoms are not under control. Everyday life is disrupted. However, care remains to be centered at home; (5) acute, symptoms or complications require hospitalization or other measures. Everyday life activities are cut back or severely curtailed; (6) crisis, a life-threatening situation that requires emergency care. Everyday life is placed on hold; (7) comeback, a return to everyday life activities, possibly with changed ability for everyday life activities; (8) downward, decline associated with increased disability and trouble controlling symptoms, requires adaptation in everyday life activities; and (9) dying, death of the patient.

Types of Information Shared

HICs provide demographic and medical information in their questions that represent their understanding of their diseases and communicate their information needs to others [32]. Demographic information included age, gender, ethnicity, weight, location, and profession. Medical information includes symptoms, medical tests, treatment, time of treatment, lifestyle, drugs, personal and family medical history, insurance, and time in hospital.

To facilitate question analysis, we developed a web-based annotation system. Two annotators coded the questions independently, with one having a medical degree. Each coder was asked to determine whether any of the categories in the coding schema was present or absent in the question. The questions were split randomly between the coders with 20% overlap to check for intercoder agreement. A comparison between the two sets of coding results showed that the intercoder agreement over the overlapping data was 87.9%. Discrepancies in the coding results were discussed and resolved among the coders, who then used the results of this discussion to review and revise the overall questions until they reached an agreement.

Results

Descriptive Statistics

Descriptive statistics for the data sets are presented in Table 1. Among the questions, 85.1% (531/624) involved single diseases and 14.9% (93/624) involved multimorbidity. Specifically, the single-disease questions accounted for approximately 50.6% (316/624) of questions from Yahoo! Answers and 49.3% (308/624) questions from WebMD Answers, respectively; multimorbid accounted for approximately 8.0% (50/624) of questions on Yahoo! Answers and 6.9% (43/624) on WebMD Answers, respectively. We performed one-way analysis of variance (ANOVA) to test the effect of disease complexity on the selected characteristics of the posted questions. The analysis was followed up with Tukey honestly significant difference test to examine the differences between different values of the variables.

Table 1. Descriptive statistics of the data sets and analysis of variance results of disease complexity.

Variable	Single (n=531; 85.1%), mean (SD)	Multimorbidity (n=93; 14.9%), mean (SD)	*F* test (*df*)	*P* value
Types of information needs				
Symptom	0.12 (0.33)	0.08 (0.27)	1.72 (1,622)	.19
Cause	0.19 (0.39)	0.14 (0.35)	1.16 (1,622)	.28
Diagnose	0.22 (0.41)	0.12 (0.32)	5.08 (1,622)	.02[a]
Treatment	0.16 (0.37)	0.26 (0.44)	4.82 (1,622)	.02[a]
Prognoses	0.13 (0.33)	0.17 (0.38)	1.31 (1,622)	.25
Drug	0.04 (0.19)	0.06 (0.25)	1.42 (1,622)	.23
Lifestyle	0.13 (0.34)	0.15 (0.36)	0.29 (1,622)	.59
Similar	0.03 (0.17)	0.05 (0.23)	1.66 (1,622)	.19
Source	0.06 (0.23)	0.09 (0.28)	1.03 (1,622)	.31
Others	0.23 (0.42)	0.28 (0.45)	1.26 (1,622)	.26
Stages of disease development				
Health stages of questions	3.92 (3.43)	6.51 (2.59)	48.02 (1,622)	<.001[b]
Chronic stages	1.04 (1.6)	2.4 (1.87)	54.01 (1,622)	<.001[b]
Type of personal information shared				
Demographic	0.21 (0.41)	0.48 (0.5)	32.24 (1,622)	<.001[b]
Medical diagnosis	0.4 (0.49)	0.58 (0.5)	11.04 (1,622)	<.001[b]
Treatment and prevention	0.25 (0.43)	0.44 (0.5)	14.55 (1,622)	<.001[b]

[a]Significant as $P<.05$.

[b]Significant as $P<.001$.

Types of Information Needs

The ANOVA results (Table 1) showed a significant effect of disease complexity on two information-seeking needs: diagnosis ($F_{1,622}=5.08$; $P=.02$) and treatment ($F_{1,622}=4.82$; $P=.02$). However, disease complexity did not have any effect on other types of information-seeking needs such as symptoms ($P=.19$), causes ($P=.28$), prognosis ($P=.25$), drug interactions ($P=.23$), lifestyle ($P=.59$), similar experiences ($P=.19$), source ($P=.31$), and others ($P=.26$).

The post-hoc comparison showed that the single-disease questions tend to include more diagnostic information (mean 0.22, SD 0.41; $P=.02$ than the multimorbid counterpart, whereas the multimorbid questions tend to contain more treatment-related information than single-disease ones (mean 0.26, SD 0.44; $P=.02$).

Stages of Disease Development

There were statistically significant differences in the health stage ($F_{1,622}=48.02$; $P<.001$) and chronic stage ($F_{1,622}=54.01$; $P<.001$) between the two levels of disease complexity. Specifically, multimorbid questions included questions more frequently at health stages of questions (mean 6.51, SD 2.59; $P<.001$) and at advanced or chronic stages of illness (mean 2.4, SD 1.87; $P<.001$).

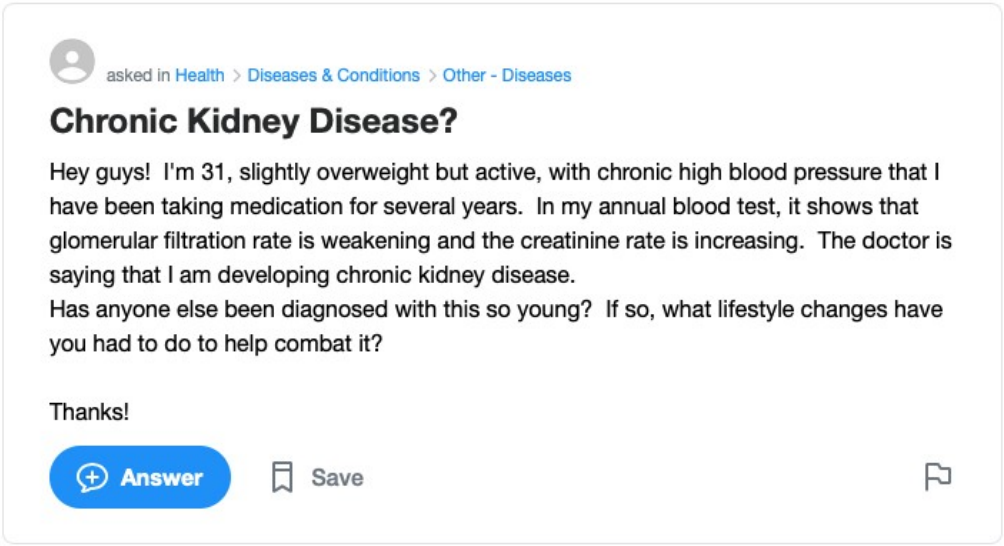
Types of Information Shared

The ANOVA results revealed the effects of disease complexity on the types of information shared, including demographic information ($F_{1,622}=32.24$; $P<.001$), medical diagnosis ($F_{1,622}=11.04$; $P<.001$), and treatment and prevention ($F_{1,622}=14.55$; $P<.001$). Specifically, questions from multimorbid HICs are more likely to provide demographic information such as age, ethnicity, and weight (mean 0.48, SD 0.5; $P<.001$) in their questions compared with single-disease HICs (mean 0.21, SD 0.41; $P<.001$). In addition, multimorbid HICs are more likely to include information relating to diagnosis (mean 0.58, SD 0.5; $P<.001$) and information relating to treatment and prevention (mean 0.44, SD 0.5; $P<.001$).

Figures 1 and 2 present two illustrations of information sharing in health questions. In the single-disease question (Figure 1), the HIC sought a possible diagnosis by sharing a list of symptoms. However, the question contained a minimal amount of demographic and medical information including medical diagnosis or information relating to treatment and prevention. In the multidisease question (Figure 2), the HIC shared some personal information, including demographics, medical diagnosis, treatment, and multimorbidity, and is seeking very specific information relating to diagnosis and lifestyle.

Figure 1. Illustrations of information seeking and sharing behavior in single-disease health questions.

Figure 2. Illustrations of information seeking and sharing behavior in multi-disease health questions.



Discussion

Principal Findings and Explanations

This research aims to investigate whether disease complexity affects information-seeking needs, stages of disease development, and the type of shared information on Q&A sites. Our empirical data analysis results revealed several significant effects.

Questions relating to single diseases were more likely to include questions about diagnosis when compared with multimorbid questions. This is not in line with medical care assumptions that underdiagnoses are a real threat to managing patients with multiple chronic conditions [34], who may not be allocated sufficient time for assessment in clinics. The results are also surprising if we consider the challenges involved in making sense of the plethora of symptoms and the complex interaction between different diseases. The reasons behind this are not clear, as there are no differences in the number of question posts asking about two related topics: symptoms and causes. Disease complexity was not found to have an effect on any other type of information needs, including symptoms, causes, prognoses, drugs, lifestyle, similar experiences, sources, and others. Potentially, patients with multimorbidity may have a better understanding of the diagnostic pathways and are content with the information they already know. In addition, the finding of this study supports that multimorbid HICs asking about treatment more frequently than single-disease HICs. This is in line with previous research assuming that people with major illnesses, such as cancer, ask mainly for advice; survivors seem to share personal narratives; and terminal patients seek acknowledgment and validation of their choices [8]. Similarly, multiple chronic conditions may require more complex management.

Another major finding is the disease stage. Multimorbid questions focused more frequently on advanced stages of disease development. We provide several alternative explanations for this finding. First, HICs at a later stage of a chronic disease may need more information to validate their choices, such as looking for detailed clinical information, which experts cannot do for various reasons. Patients might understand that care is patient centered, and it is common for treatments to differ, especially across different states or countries. In addition, patients at later stages may require more general information, such as lifestyle, for which peers can serve as great information sources.

Disease complexity was found to play a major role in determining the types of information shared on web-based Q&A sites. In particular, the multimorbid HICs included more demographic and medical information in their questions, which included information related to diagnosis and information relating to treatment and prevention. Although the management of chronic disease is a highly collaborative process between patients and providers, the work that patients must take on

during the various stages of chronic disease progression is immense [35,36]. Multimorbid HICs ask for additional information about management and treatments of their diseases, further highlighting their active involvement in health self-management.

Contribution and Design Implications

This study makes several novel scientific contributions to health consumer informatics. To the best of our knowledge, this is the first study to empirically investigate the effects of disease complexity on the types of information shared in two Q&A sites. In addition, this is the first study to examine the disease stages of HICs with respect to disease complexity. Overall, this study provides a comprehensive examination of a wide variety of question characteristics, which is unique among studies on web-based health Q&A sites.

This study used mixed methods that combine qualitative and quantitative methods to understand health information–seeking behavior in web-based Q&A sites. In addition, the study considered two common types of Q&A sites, community based and expert curated, for analyzing health information–seeking behavior.

Unlike previous studies that view information sharing and information seeking as two conflicting goals [37-39], the findings of this study suggest that information sharing can facilitate information seeking in online health communities. Previous studies have focused more on individual motivations and intentions to share information but less on the actual sharing behaviors and the content of sharing [40]. In addition, we extended previous theories of information-seeking behavior by accounting for the dynamics of information needs; in other words, the process of information seeking can vary with the stages of disease development.

Our research findings have several implications for improving web-based Q&A sites. The information needs of HICs with chronic diseases may also change over time as their disease evolves, such as substantial disruptions to their everyday lives. As a result, effective support from peers who share similar characteristics and experiences would be very helpful. It is worth noting that HICs with a chronic disease grow their knowledge as they continuously manage their conditions and take more responsibility for their illnesses [33]. However, there is a lack of research investigating how the use of web-based information evolves as HICs gain experience moving through a chronic disease trajectory. How to design systems for HICs with chronic diseases that can not only support them in managing their own evolving health conditions and related knowledge but also enable them to help others with similar conditions are important questions for future research.

This study provides strong evidence that multimorbid HICs share both demographic and medical information when they seek health information on the web. Although web-based Q&A sites encourage the exchange of a significant amount of health information, these sites can benefit from improving the organization of information for community reuse. For instance, these sites could better support HICs with specific information needs by organizing the questions based on specific types of information needs, such as diagnosis, treatment, and side effects, and by encouraging them to share their demographic and medical information without compromising their personal privacy.

Limitations and Future Directions

Our study has several limitations. As far as the data sources are concerned, we collected health questions from two types of Q&A sites. Thus, caution should be exercised when generalizing the findings to other Q&A sites. For instance, Twitter has been used for Q & A in the health context. It would help enrich the Q&A literature to build a tweets data set on health Q&A and use it to validate the findings of this study. In addition, multiple disease questions had a much smaller proportion as opposed to a single-disease question in this study. Furthermore, we have chosen to focus on health topics on the kidney disease. Although the findings of this study are expected to be extensible to other chronic diseases, they still require empirical validations. In addition to the health questions, which is the focal point of the analysis in this study, the Q&A sites also provide other types of information such as user profiles, comments, answers, and ratings. Integrating information from these multiple dimensions is expected to achieve a deeper understanding of the web-based behavior of HICs. On the other hand, the disclosures of increasing amount of personal health information may raise privacy concerns. Thus, how HIC trade-off between information needs and information disclosure is an interesting question for future research. The qualitative approach used in this study helps uncover the contextual characteristics of questions, which however is difficult to scale. Text mining has the potential to address the limitation by automating this process. The data generated through this study can serve as the training data for building text-mining models.

Conclusions

Multiple disease or multimorbidity questions seem to play a major role in the stages of disease development and types of information shared, highlighting a deeper understanding of the complexities of their conditions. Regarding the types of information needs, multimorbidity has a minor implication related to treatment. It is also the case of single-disease questions that seem to be relevant only for types of information needs in terms of diagnosis. This study is a valuable first step in investigating the effects of multimorbidity on different types of information shared in two Q&A sites. The findings present implications for designing web-based Q&A sites to better support health information seeking.

Authors' Contributions

AA developed the study, analyzed the data, and wrote the manuscript. LZ reviewed the manuscript and made the necessary edits.

Conflicts of Interest

None declared.

References

1. Fox S. The Social Life of Health Information. Pew Research Center. 2014. URL: http://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/ [accessed 2019-12-06] [WebCite Cache ID http://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/]
2. Shah C, Oh S, Oh JS. Research agenda for social Q&A. Library & Information Science Research 2009 Dec;31(4):205-209. [doi: 10.1016/j.lisr.2009.07.006]
3. Oh S. Social Q&A. Social Information Access 2018;10100:75-107. [doi: 10.1007/978-3-319-90092-6_3]
4. Emerson RM. Social exchange theory. Annu Rev Sociol 1976 Aug;2(1):335-362. [doi: 10.1146/annurev.so.02.080176.002003]
5. Cross R, Rice R, Parker A. Information seeking in social context: structural influences and receipt of information benefits. IEEE Trans Syst 2019;31(4):438-448. [doi: 10.1109/5326.983927]
6. Xiao N, Sharman R, Rao H, Upadhyaya S. Factors influencing online health information search: An empirical analysis of a national cancer-related survey. Decision Support Systems 2014 Jan;57:417-427. [doi: 10.1016/j.dss.2012.10.047]
7. Oh S. The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments. J Am Soc Inf Sci 2011 Nov 1;63(3):543-557. [doi: 10.1002/asi.21676]
8. Oh S, Zhang Y, Park M. Cancer information seeking in social question and answer services: identifying health-related topics in cancer questions on Yahoo! Answers. Information Research 2016;21(3):10-16.
9. Park MS, He Z, Chen Z, Oh S, Bian J. Consumers' Use of UMLS Concepts on Social Media: Diabetes-Related Textual Data Analysis in Blog and Social Q&A Sites. JMIR Med Inform 2016 Nov 24;4(4):e41 [FREE Full text] [doi: 10.2196/medinform.5748] [Medline: 27884812]
10. Doing-Harris KM, Zeng-Treitler Q. Computer-assisted update of a consumer health vocabulary through mining of social network data. J Med Internet Res 2011 May 17;13(2):e37 [FREE Full text] [doi: 10.2196/jmir.1636] [Medline: 21586386]
11. Oh JS, He D, Jeng W, Mattern E, Bowler L. Linguistic characteristics of eating disorder questions on Yahoo! Answers - content, style, and emotion. Proc Am Soc Info Sci Tech 2014 May 8;50(1):1-10. [doi: 10.1002/meet.14505001068]
12. Johnson JD, Meischke H. Women's preferences for cancer information from specific communication channels. American Behavioral Scientist 2016 Jul 27;34(6):742-755. [doi: 10.1177/0002764291034006010]
13. Hesse BW, Arora NK, Burke BE, Finney RL. Information support for cancer survivors. Cancer 2008 Jun 1;112(11 Suppl):2529-2540 [FREE Full text] [doi: 10.1002/cncr.23445] [Medline: 18428201]
14. Boyd CM, Darer J, Boult C, Fried LP, Boult L, Wu AW. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance. J Am Med Assoc 2005 Aug 10;294(6):716-724. [doi: 10.1001/jama.294.6.716] [Medline: 16091574]
15. Fraser SDS, Taal MW. Multimorbidity in people with chronic kidney disease: implications for outcomes and treatment. Curr Opin Nephrol Hypertens 2016 Nov;25(6):465-472. [doi: 10.1097/MNH.0000000000000270] [Medline: 27490909]
16. Sultan M, Kuluski K, McIsaac WJ, Cafazzo JA, Seto E. Turning challenges into design principles: telemonitoring systems for patients with multiple chronic conditions. Health Informatics J 2019 Dec;25(4):1188-1200. [doi: 10.1177/1460458217749882] [Medline: 29320911]
17. Alasmari A, Zhou L. How multimorbid health information consumers interact in an online community Q&A platform. Int J Med Inform 2019 Nov;131:103958. [doi: 10.1016/j.ijmedinf.2019.103958] [Medline: 31521012]
18. Doyle J, Murphy E, Smith S, Hannigan C, Kuiper J, Jacobs A. Addressing medication management for older people with multimorbidities: a multi-stakeholder approach. New York, New York, USA: ACM Press; 2017 Presented at: PervasiveHealth '17: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare; May, 2017; Barcelona Spain p. 78-87. [doi: 10.1145/3154862.3154883]
19. Ongwere T, Cantor G, Martin S, Shih P, Clawson J, Connelly K. Design hotspots for care of discordant chronic comorbidities: patients' perspectives. New York, New York, USA: ACM Press; 2018 Presented at: NordiCHI '18: Proceedings of the 10th Nordic Conference on Human-Computer Interaction; September, 2018; Oslo Norway p. 571-583. [doi: 10.1145/3240167.3240221]
20. De Nicola L, Zoccali C. Chronic kidney disease prevalence in the general population: heterogeneity and concerns. Nephrol Dial Transplant 2016 Mar 29;31(3):331-335. [doi: 10.1093/ndt/gfv427] [Medline: 26908831]
21. Coyne D. Management of Chronic Kidney. CKD Medscape CME Expert Column Series 2011(3):1-13.
22. Honeycutt AA, Segel JE, Zhuo X, Hoerger TJ, Imai K, Williams D. Medical costs of CKD in the Medicare population. J Am Soc Nephrol 2013 Sep;24(9):1478-1483 [FREE Full text] [doi: 10.1681/ASN.2012040392] [Medline: 23907508]
23. Weder N, Zhang H, Jensen K, Yang BZ, Simen A, Jackowski A, et al. Child abuse, depression, and methylation in genes involved with stress, neural plasticity, and brain circuitry. J Am Acad Child Adolesc Psychiatry 2014 Apr;53(4):417-24.e5 [FREE Full text] [doi: 10.1016/j.jaac.2013.12.025] [Medline: 24655651]
24. Uhlig K, Boyd C. Guidelines for the older adult with CKD. Am J Kidney Dis 2011 Aug;58(2):162-165. [doi: 10.1053/j.ajkd.2011.05.001] [Medline: 21664736]

25. Levey AS, Atkins R, Coresh J, Cohen EP, Collins AJ, Eckardt K, et al. Chronic kidney disease as a global public health problem: approaches and initiatives - a position statement from Kidney Disease Improving Global Outcomes. Kidney Int 2007 Aug;72(3):247-259. [doi: 10.1038/sj.ki.5002343] [Medline: 17568785]
26. Fraser SD, Roderick PJ, May CR, McIntyre N, McIntyre C, Fluck RJ, et al. The burden of comorbidity in people with chronic kidney disease stage 3: a cohort study. BMC Nephrol 2015 Dec 1;16:193 [FREE Full text] [doi: 10.1186/s12882-015-0189-z] [Medline: 26620131]
27. Bramall AN, Bernstein M. Improving information provision for neurosurgical patients: a qualitative study. Can J Neurol Sci 2014 Jan 23;41(1):66-73. [doi: 10.1017/s0317167100016280] [Medline: 24384340]
28. Nowrouzi B, Gohar B, Nowrouzi-Kia B, Garbaczewska M, Brewster K. An examination of scope, completeness, credibility, and readability of health, medical, and nutritional information on the internet: a comparative study of wikipedia, webmd, and the mayo clinic websites. Canadian Journal of Diabetes 2015 Apr;39:S71. [doi: 10.1016/j.jcjd.2015.01.267]
29. The Top 500 Sites on the Web. Alexa. URL: https://www.alexa.com/topsites/category/Top/Reference [accessed 2020-05-01] [WebCite Cache ID https://www.alexa.com/topsites/category/Top/Reference]
30. Kidney Problem Symptoms, Causes, and Types. American Kidney Fund. URL: https://www.kidneyfund.org/kidney-disease/kidney-problems/ [accessed 2020-05-07]
31. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
32. Zhang Y. Contextualizing consumer health information searching: an analysis of questions in a social Q&A community. In: Proceedings of the 1st ACM International Health Informatics Symposium. 2010 Presented at: IHI'10; May 2-4, 2010; Arlington Virginia USA. [doi: 10.1145/1882992.1883023]
33. Corbin JM. The corbin and strauss chronic illness trajectory model: an update. Sch Inq Nurs Pract 1998;12(1):33-41. [Medline: 9805470]
34. Hausmann D, Kiesel V, Zimmerli L, Schlatter N, von Gunten A, Wattinger N, et al. Sensitivity for multimorbidity: The role of diagnostic uncertainty of physicians when evaluating multimorbid video case-based vignettes. In: PLoS One. Sensitivity for multimorbidity: The role of diagnostic uncertainty of physicians when evaluating multimorbid video case-based vignettes. Martínez Molina A. editor. PLoS One. Public Library of Science; 2019:e0215049.
35. Corbin J, Strauss A. Managing chronic illness at home: Three lines of work. Qualitative Sociology 1985;8(3):224-247. [doi: 10.1007/BF00989485]
36. Holman H, Lorig K. Patient self-management: a key to effectiveness and efficiency in care of chronic disease. Public Health Rep 2004 Aug 02;119(3):239-243 [FREE Full text] [doi: 10.1016/j.phr.2004.04.002] [Medline: 15158102]
37. Moncur W, Reiter E, Masthoff J, Carmichael A. Modeling the socially intelligent communication of health information to a patient's personal social network. IEEE Trans Inform Technol Biomed 2010 Mar;14(2):319-325. [doi: 10.1109/titb.2009.2035361]
38. Skeels M. Sharing by Design: Understanding and Supporting Personal HealthInformation Sharing and Collaboration within Social Networks Internet. In: University of Washington. NA: ProQuest Dissertations Publishing; 2010:2010.
39. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun 14;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
40. Zhang X, Liu S, Chen X, Gong Y. Social capital, motivations, and knowledge sharing intention in health Q&A communities. Management Decision 2017;55(7):1536-1557.

Abbreviations

ANOVA: analysis of variance
HIC: health information consumer
Q&A: question and answer

Edited by R Kukafka; submitted 07.07.20; peer-reviewed by E Alanazi, S Song; comments to author 23.08.20; revised version received 08.10.20; accepted 10.01.21; published 24.03.21.

Please cite as:
Alasmari A, Zhou L
Share to Seek: The Effects of Disease Complexity on Health Information–Seeking Behavior
J Med Internet Res 2021;23(3):e21642
URL: https://www.jmir.org/2021/3/e21642
doi:10.2196/21642
PMID:33759803

Original Paper

Internet-Specific Epistemic Beliefs in Medicine and Intention to Use Evidence-Based Online Medical Databases Among Health Care Professionals: Cross-sectional Survey

Yen-Lin Chiu[1], PhD; Yu-Chen Lee[2,3], MD, PhD; Chin-Chung Tsai[4,5], PhD

[1]Department and Graduate Institute of Medical Education and Bioethics, National Taiwan University College of Medicine, Taipei, Taiwan

[2]Department of Chinese Medicine, China Medical University Hospital, Taichung, Taiwan

[3]Graduate Institute of Acupuncture Science, China Medical University, Taichung, Taiwan

[4]Program of Learning Sciences, National Taiwan Normal University, Taipei, Taiwan

[5]Institute for Research Excellence in Learning Sciences, National Taiwan Normal University, Taipei, Taiwan

Corresponding Author:
Chin-Chung Tsai, PhD
Program of Learning Sciences
National Taiwan Normal University
No 162, Sec 1, He-Ping E Rd
Taipei, 10610
Taiwan
Phone: 886 2 7749 5179
Email: tsaicc@ntnu.edu.tw

Abstract

Background: Evidence-based medicine has been regarded as a prerequisite for ensuring health care quality. The increase in health care professionals' adoption of web-based medical information and the lack of awareness of alternative access to evidence-based online resources suggest the need for an investigation of their information-searching behaviors of using evidence-based online medical databases.

Objective: The main purposes of this study were to (1) modify and validate the internet-specific epistemic beliefs in medicine (ISEBM) questionnaire and (2) explore the associations between health care professionals' demographics, ISEBM, and intention to use evidence-based online medical databases for clinical practice.

Methods: Health care professionals in a university-affiliated teaching hospital were surveyed using the ISEBM questionnaire. The partial least squares-structural equation modeling was conducted to analyze the reliability and validity of ISEBM. Furthermore, the structural model was analyzed to examine the possible linkages between health professionals' demographics, ISEBM, and intention to utilize the evidence-based online medical databases for clinical practice.

Results: A total of 273 health care professionals with clinical working experience were surveyed. The results of the measurement model analysis indicated that all items had significant loadings ranging from 0.71 to 0.92 with satisfactory composite reliability values ranging from 0.87 to 0.94 and average variance explained values ranging from 0.70 to 0.84. The results of the structural relationship analysis revealed that the source of internet-based medical knowledge (path coefficient –0.26, P=.01) and justification of internet-based knowing in medicine (path coefficient 0.21, P=.001) were correlated with the intention to use evidence-based online medical databases. However, certainty and simplicity of internet-based medical knowledge were not. In addition, gender (path coefficient 0.12, P=.04) and academic degree (path coefficient 0.15, P=.004) were associated with intention to use evidence-based online medical databases for clinical practice.

Conclusions: Advancing health care professionals' ISEBM regarding source and justification may encourage them to retrieve valid medical information through evidence-based medical databases. Moreover, providing support for specific health care professionals (ie, females, without a master's degree) may promote their intention to use certain databases for clinical practice.

(J Med Internet Res 2021;23(3):e20030) doi:10.2196/20030

KEYWORDS

evidence-based medicine (EBM); health care professionals; internet-specific epistemic beliefs; medical informatics

Introduction

Evidence-Based Medicine

Evidence-based medicine, defined as an integration of the best available external evidence, individual clinical expertise, and patient preferences in making optimal decisions for patient care, remains a major concern for health care professionals, public health practitioners, and medical educators [1-3]. Applying the most efficacious treatment integrated with evidence-based medicine will maximize the quality and quantity of life of individual patients [1]. Practicing evidence-based medicine may promote rapid updating of knowledge, improve expected patient care, and have positive effects on doctor-patient relationships [4-6].

As the internet provides a broad range of web-based medical information for health care professionals in their daily practice, their web-based medical information–searching behaviors have become an issue of great interest [7,8]. The internet has become the most frequently used resource for obtaining medical information in clinical practice [9], while also being considered as an evidence-based medical tool [10]. Evidence-based online medical databases may support convenient access for health care professionals to search for and retrieve evidence-based medical information in clinical contexts [3,11,12]. Using evidence-based online medical databases to improve patient care is regarded as a legitimate clinical role of health care professionals [11]. Information literacy, that is, awareness and evaluation of the evidence sources, is also regarded as a necessary competency of health professionals in evidence-based practice [13].

Information-Searching Behaviors

In spite of the help that web-based medical information provides in clinical decision making, its quality and credibility is an issue of concern, thereby suggesting the need to browse medical databases such as PubMed to cross-check web-based information and retrieve the best evidence [14]. Retrieving and applying information from evidence-based biomedical websites in clinical contexts is essential for health care professionals to improve patient care [2,15]. Most nurses and physicians (93%) who used evidence-based online medical databases believed that their use may help improve patient care [11]. Physicians with high usage of authoritative information sources, including scientific web-based databases such as PubMed and scientific digital journals, possess high fulfilment in meeting information needs and perceive themselves as having medical practice competency [7]. Clinicians' use of evidence-based electronic knowledge resources has positive effects on their behaviors and on patients' outcomes [16].

Although a large proportion of nurses use Google (80.2%), thereby making Google a regularly utilized electronic information resource, they seldom use web-based medical databases such as MEDLINE (19.8%) and Cochrane (1.1%) [9]. Moreover, there is a tendency for physicians to use nonauthoritative information sources such as Wikipedia, Facebook groups, and YouTube when retrieving web-based medical information [7]. Even though a majority of the health care professionals positively value the promotion of evidence-based medicine, they may not be aware of evidence-based medical information available online and may lack the ability to access validated evidence-based resources [3,17]. Further, medical students adopt less sophisticated information-searching strategies than their counterparts in general universities, thereby suggesting that additional training in searching for information on the internet is warranted [18].

Regarding the demographics of evidence-based online medical database users, the more senior the health care professionals are, the less likely they are to access evidence-based medical information on the internet [17]. Moreover, nurses require training in evidence-based practice, in particular, senior nurses with work experience of more than 5 years [19]. The age and educational level of physicians and nurses are also considered as important factors in predicting their usage of evidence-based online medical databases [11,20]. Regarding gender, male physicians are more likely than female physicians to employ web portals to search for medical information [20]. Thus, there are associations between the usage of internet-based medical resources and user characteristics such as gender, work experience, and academic degree [17,19,20].

Epistemic Beliefs and Medical Information Searching on the Internet

Epistemic beliefs, a construct with multiple dimensions relating to the nature of knowledge and the way of knowing, have been defined as personal cognitions, including certainty of knowledge, simplicity of knowledge, source of knowledge, and justification for knowing [21]. Understanding how medical learners perceive and acquire medical knowledge (ie, their epistemic beliefs about medicine) has been considered an important issue in medical education [22,23]. Examining the relationship between epistemic beliefs and medical learning has potential for improving medical education [23]. Additionally, epistemic beliefs are regarded as an influential factor in searching on the internet [24,25]. While solving ill-structured problems through web-based information searching, the activation of epistemic beliefs may help learners who possess complex epistemic beliefs to use advanced search strategies, evaluate information quality, and search for alternative views [24]. Research on epistemic beliefs has focused on its role in processing diverse and conflicting information on the internet [26-28].

With respect to medical information searching, epistemic beliefs have also been defined as a determinant factor in dealing with controversial medical information and making health-related decisions [25,29-31]. Research on the navigation behaviors of adults without a university education showed that while searching for health-related issues on the internet, beliefs regarding justification for multiple resources and the reliability of web-based knowledge were positively related to the time spent on the objective webpages and recommended postsearch health decisions [31]. However, the lack of research on the influences of epistemic beliefs on health care professionals' medical information–searching behaviors needs to be noted and explored.

Owing to the context-sensitive nature of personal epistemology, it has been suggested that beliefs about knowledge and knowing should be investigated in a specific context (eg, web-based

searching) and measured with a context-specific instrument [27,28]. To evaluate epistemic beliefs in an internet-based environment, the construct and measurement of internet-specific epistemic beliefs have been developed to assess individual beliefs regarding internet-based knowledge and knowing [32]. In addition, the construct validity of internet-specific epistemic beliefs measurement has been rigorously examined and its relationships with web-based information search activities have been extensively explored [33-35]. Based on the four-dimension theory of personal epistemology developed by Hofer and Pintrich [21], the dimensions of the internet-specific epistemic beliefs questionnaire (ISEQ) were originally constructed by Bråten and his colleagues [32]. Additionally, the ISEQ was utilized to examine the roles of internet-specific beliefs in web-based health information searching behaviors and web search activities on a medical issue [31,35].

The Chinese version of the internet-specific epistemic beliefs questionnaire (C-ISEQ), adapted from ISEQ and translated into Chinese, was utilized to measure university students' beliefs regarding internet-based knowledge and knowing in Taiwan [33]. The C-ISEQ has been rigorously modified and validated as having acceptable reliability and appropriate validity, denoting a four-dimension model with 12 items related to beliefs in certainty of internet-based knowledge, simplicity of internet-based knowledge, source of internet-based knowledge, and justification for internet-based knowing [33,34]. In addition, the C-ISEQ has been utilized to examine the relationships between internet-specific epistemic beliefs and web-based information-searching behaviors for course-related questions [33,36].

There are disciplinary differences in epistemic beliefs across domains [37,38]. For example, students of science and psychology have different views on certainty, source, justification, and truth of knowledge [37]. Furthermore, the domain specificity of epistemic beliefs has been referred to and explored, thereby showing the influential role of scientific epistemic beliefs in learning science [39,40]. With respect to medical education, medical students' medicine-related epistemic beliefs may be related to their approach to learning medicine [22]. Presumably, when investigating web-based search behaviors in the context relating to a specific domain, there is a need to consider individuals' epistemic beliefs in terms of their context specificity (eg, internet-specific) and their domain specificity (eg, medicine-related) simultaneously. Owing to a deficiency of epistemic beliefs measurement for both internet-based and medicine-related contexts, there is a need to develop an instrument to measure the internet-specific epistemic beliefs in medicine (ISEBM).

Research Objectives

Clinicians' lack of intention to access evidence-based medical information sources remains a challenge [13,17]. The behaviors of health care professionals in searching web-based evidenced resources and evaluating the reliability of the retrieved information have become an issue of concern and have been increasingly examined [3,12,13,41]. Based on the help of advanced epistemic beliefs in skilled search strategies and valid information evaluation [28,42], there may be a presumable relationship between health care professionals' epistemic beliefs and their intention to search for evidence-based medical information in web-based biomedical databases. Considering the context, as well as the domain specificity of web-based medical information searching [32,38], the first purpose of this study was to modify and validate a measurement to assess internet-specific and medicine-related epistemic beliefs, which have not been previously researched. Previous studies targeted laypeople rather than health care professionals in exploring the relations between internet-specific epistemic beliefs and web-based medical information searching activities [31,35]. To the best of our knowledge, no research has focused on health care professionals' epistemic beliefs in medicine and their relationship with evidence-based medicine. Therefore, the second purpose of this study was to explore the relationship between health care professionals' ISEBM and their intention to use evidence-based online medical databases such as MEDLINE and Cochrane while retrieving medical information on the internet. In sum, the main research questions of this study are as follows:

1. Is the instrument utilized in this study valid and reliable for measuring health care professionals' beliefs regarding internet-based medical knowledge and knowing?

2. What are the relationships between health care professionals' internet-specific beliefs about medicine and their intention to utilize evidence-based online medical databases?

Methods

Recruitment

In 1 university-affiliated teaching hospital in Taiwan, this study purposefully recruited health care professionals with work experience of more than 6 months. All the participants in this study voluntarily responded to the survey. Before answering the questionnaire, they read the cover statement that stated the confidential nature of this survey and were informed that they were free to withdraw from the study. The return of the finished questionnaire was regarded as their consent to participate.

Measures

ISEBM

Considering the same language and the cultural context, the C-ISEQ was adopted from the previous work of Chiu et al to develop the major measure of this study, namely, the ISEBM questionnaire [33]. In addition, the items of internet-specific epistemic beliefs developed in the work of Bråten et al were retrieved and included in the ISEBM questionnaire [32]. According to the suggested process, the questionnaire development was conducted in the following steps [43]. First, the construct of internet-specific beliefs was defined and discussed before the development of the questionnaire to ensure the content validity of the questionnaire. Then, the wording of each item of the ISEBM was carefully modified to assess individuals' internet-specific epistemic beliefs about web-based medical knowledge and knowing. Next, the ISEBM was checked by experts to confirm the face validity. Finally, the statistical estimates of reliability and validity were calculated via the use

of the structural equation modeling (SEM) technique to examine the validation of the ISEBM.

Originating from the theoretical constructs of C-ISEQ [33], the ISEBM questionnaire was constructed with 4 dimensions, namely ISEBM-CE (certainty), ISEBM-SP (simplicity), ISEBM-SO (source), and ISEBM-JU (justification). With regard to the domain-specific nature of epistemic beliefs, the specificity of domain knowledge should be considered while developing the measurement of epistemic beliefs in certain domains [38]. To take the specificity of medicine into consideration, a total of 18 items of internet-specific epistemic beliefs retrieved from prior studies [32,33] were revised with wordings to specify the internet-specific epistemic beliefs relating to web-based searching for medical information. As suggested by researchers, the face validity should be initially established when using borrowed measurements to develop a new instrument [43]. Prior to the analysis of construct validity through the statistical technique, 2 experts in medicine and information education evaluated the content and meaning of each item and asserted the face validity of ISEBM.

The four-dimension ISEBM consists of 18 items measured with a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The higher scores of certainty, simplicity, and source dimension represent the more naïve beliefs in the internet-based knowledge, that is, the respondents are more likely to believe that the knowledge retrieved from the internet is certain, simple, and accurate. On the contrary, higher scores on the justification dimension denote that the respondents possess the sophisticated belief that the internet-based knowledge claims should be carefully justified. The concepts of each dimension are illustrated in the following paragraph.

The certainty of internet-based knowledge in medicine measures beliefs that the medical knowledge found on the internet is certain and true. Respondents with high scores on certainty, indicating naïve beliefs, are less likely to doubt the certainty of medical knowledge found on the internet. Simplicity of internet-based knowledge in medicine assesses beliefs that the internet-based medical knowledge is simple and specific. A high score on simplicity implies the naïve view that the internet contains simple and detailed medical knowledge. Source of internet-based knowledge in medicine evaluates beliefs that the internet is a good source that offers correct and essential medical knowledge. Respondents with high scores, showing naïve views on source, are more likely to believe that the internet contains good and accurate medical knowledge. Justification for internet-based knowing in medicine examines the extent to which respondents believe that the medical claims on the internet should be evaluated and justified. Respondents with high scores on justification, demonstrating sophisticated epistemic views, believe that the internet-based medical claims should be critically evaluated against other sources.

Demographic Variables

In addition to the ISEBM questionnaire, the participants' gender, years of working experience, and academic degree were included in the structural model of SEM and were regarded as independent variables. The female participants were treated as the reference group and coded as 0, while the male participants were coded as 1. Years of work experience was treated as a continuous variable, indicating the actual years of clinical work experience. Academic degree was coded as a 2-type categorical variable, including bachelor's degree level (coded as 1) and master's degree level (coded as 2). Regarding the dependent variable, intention to use evidence-based online medical databases was a single item measured with a 5-point Likert scale ranging from 1 (impossible) to 5 (extremely possible). It was retrieved and modified from prior studies, which investigated health care professionals' willingness to perform web-based learning activities [44]. While answering this item, the participants indicated the extent to which they may employ the evidence-based online medical databases such as MEDLINE, Cochrane, and UpToDate when retrieving web-based medical information to answer medical problems in clinical contexts.

Statistical Analysis

With respect to the participants' demographics, descriptive analyses, cross-tabulation analysis, and one-sided *t* test analyses were conducted using the software SPSS version 22 (IBM Corp). Following the guidelines of the SEM analytic approach recommended by Hair et al [45], the measurement model of the ISEBM instrument and the structural model of the research hypotheses were examined via partial least squares-structural equation modeling (PLS-SEM) analysis. The PLS-SEM model was sequentially analyzed and interpreted in 2 stages. In the first stage, the factor loadings, composite reliability, average variance explained, and the Fornell-Lacker criterion [46] for the measurement model of the ISEBM instrument were evaluated to test the reliability and validity of the instrument. Following the two-step procedure recommendation, some measure items with inadequate estimates were deleted to reach an acceptable model fit [47]. In the second stage, the structural model (ie, path correlation analysis) was assessed to examine the relations among participants' demographics (ie, work experience, gender and academic degree), ISEBM, and intention to use evidence-based online medical databases. The software SmartPLS3 was utilized to perform the PLS-SEM analyses. *P* values less than .05 were regarded as significant loadings and statistically significant relationships between variables.

Results

Participants

This study adopted the paper-and-pencil survey approach. The data were collected from March to June 2018. After deleting 4 cases with major missing values, the data from a sample of 273 health care professionals with clinical work experience of more than 6 months in one university-affiliated teaching hospital was employed in the following calculation. The group comprised 84 physicians, 45 nurses, 57 pharmacists, 63 therapists, 18 medical technologists, and 6 nutritionists. Among them, 172 (63.0%) were females and 101 (36.9%) were males. Their average age was 29.57 years (range 20-65 years) and their average work experience was 4.60 years (range 0.5-37 years). Regarding their academic degrees, 50 participants possessed a master's degree, while 223 held a bachelor's degree. All participants included in this study voluntarily responded to the survey. Moreover, informed consent was obtained from the

participants. With respect to further analysis of the participant data, please refer to the statistical analyses reported in Multimedia Appendix 1.

PLS-SEM Analysis of the Measurement Model

After deleting 6 items with loadings smaller than 0.5 and 1 item with a loading higher than 0.95, the 4-dimension model containing 11 items with significant loadings ranging from 0.72 to 0.92 was identified as a reasonable measurement model. The composite reliability values ranged from 0.89 to 0.94, indicating that the reliability was fairly good (larger than 0.7). Further, the average extracted values ranged from 0.70 to 0.84, suggesting acceptable convergent validity (larger than 0.5) [48]. Table 1 represents the details of the measurement items, factor loadings, reliability, and validity indices. In addition, as shown in Table 2, based on the Fornell-Lacker criterion, the square root of average variance explained for each factor was higher than the corresponding interfactor correlations ranging from 0.02 to 0.62, showing evidence of adequate discriminant validity [46,48].

Table 1. Results of loadings, reliability, and convergent validity analysis.

Items, subitems	Loading	Composite reliability	Average variance explained	Rho value	α value
ISEBM-CE[a]		0.89	0.81	0.77	.76
ICE1. I could find accurate answers to medical problems on the internet.	0.88				
ICE2. I am most confident that I have understood medical problems when I have used the internet as a source of medical information.	0.91				
ISEBM-SP[b]		0.94	0.83	0.90	.90
ISP1. The internet provides abundant details about medical topics.	0.90				
ISP2. The internet offers simple and specific knowledge regarding medical topics.	0.92				
ISP3. The internet includes a lot of specific information related to medical issues.	0.91				
ISEBM-SO[c]		0.87	0.70	0.87	.82
ISO1. Most medical information can be found on the internet.	0.71				
ISO2. The internet involves various sources, which provide the correct answers to medical questions.	0.87				
ISO3. The internet contains information sources offering most medical knowledge.	0.91				
ISEBM-JU[d]		0.94	0.84	0.93	.91
IJU1. I would compare information from various sources to evaluate the trustworthiness of medical knowledge retrieved from the internet.	0.92				
IJU2. I would judge the logicality of the medical knowledge that I find on the internet.	0.91				
IJU3. For the same topic, I would check more sources to evaluate medical knowledge available on the internet.	0.92				

[a]ISEBM-CE: certainty of internet-specific epistemic beliefs in medicine.

[b]ISEBM-SP: simplicity of internet-specific epistemic beliefs in medicine

[c]ISEBM-SO: source of internet-specific epistemic beliefs in medicine.

[d]ISEBM-JU: justification of internet-specific epistemic beliefs in medicine.

Table 2. Results of the discriminant validity analysis.[a]

Factors	ISEBM-CE[b]	ISEBM-SP[c]	ISEBM-SO[d]	ISEBM-JU[e]
ISEBM-CE	*0.90*	—[f]	—	—
ISEBM-SP	0.41	*0.91*	—	—
ISEBM-SO	0.62	0.39	*0.84*	—
ISEBM-JU	0.02	0.49	0.18	*0.92*

[a]The correlations between factors are below the diagonal, while the square root values for average variance explained estimates (in italics) are presented on the diagonal.

[b]ISEBM-CE: certainty of internet-specific epistemic beliefs in medicine.

[c]ISEBM-SP: simplicity of internet-specific epistemic beliefs in medicine.

[d]ISEBM-SO: source of internet-specific epistemic beliefs in medicine.

[e]ISEBM-JU: justification of internet-specific epistemic beliefs in medicine.

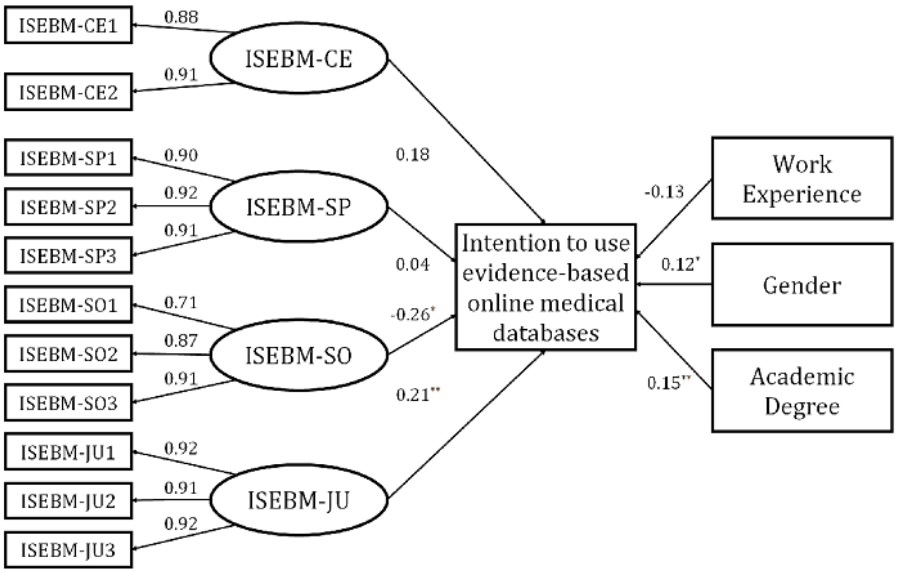
[f]Not applicable.

Path Correlations of the Structural Model

The structural model, combined with the measurement model of ISEBM, intention to use evidence-based medical databases, and demographic variables, including years of work experience, gender, and academic degree, was analyzed with PLS-SEM to evaluate the path coefficients between the variables. The values of variance inflation factor for independent variables ranged from 1.03 to 1.87, showing that there was no problem of collinearity [48]. In addition, the PLS-SEM results showed reasonable model fitness with root-mean-square error of approximation (RMSEA) of 0.071 [45].

The path coefficients are presented in Figure 1. The source of internet-based knowledge in medicine has a negative correlation (path coefficient –0.26, *P*=.01) between intention to use evidence-based online medical databases, while justification for internet-based knowing in medicine has a positive correlation (path coefficient 0.21, *P*=.001) with such intention. Regarding the demographics, gender (male) and academic degree (master's degree) were positively correlated to such intention with coefficients of 0.12 (*P*=.04) and 0.15 (*P*=.004), respectively. In all, the R^2 value for intention was 0.13, while the adjusted R^2 value was 0.10.

Figure 1. The partial least squares-structural equation modeling results for the measurement model and structural model. Intention indicates health care professionals' intention to utilize the evidence-based online medical databases. With respect to gender, male is coded as 1 while female is coded as 0. Regarding academic degree, master's degree is coded as 2 while bachelor's degree is coded as 1. *P<.05, **P<.01. ISEBM-CE: certainty of internet-specific epistemic beliefs in medicine; ISEBM-SP: simplicity of internet-specific epistemic beliefs in medicine; ISEBM-SO: source of internet-specific epistemic beliefs in medicine; ISEBM-JU: justification of internet-specific epistemic beliefs in medicine.



Discussion

Principal Findings

This study advances the understanding of epistemic beliefs in web-based information-searching contexts and explored the role of internet-specific epistemic beliefs in utilizing evidence-based online medical databases. With respect to the medical issue in web-based searching contexts, there have been a small number of studies that have investigated the internet-specific beliefs of laypersons rather than those of medical professionals [31,33,49]. Due to the widespread

utilization of the internet for searching for and locating medical knowledge in daily clinical practice [8,20], it is important to understand health care professionals' internet-specific epistemic beliefs, which play an influential role in their web-based information-searching behavior [24,26]. Assisting them with effective usage of web-based resources is important to improve patient care [8].

ISEBM and Intention to Use Evidence-Based Online Medical Databases

Individuals with sophisticated epistemic beliefs (rather than naïve beliefs) are more likely to use advanced information search strategies to evaluate the quality of web information by checking alternative sources [24]. In addition, beliefs regarding justification for multiple sources are positively related to deep web search activities [31]. As expected, the results of SEM with path analysis showed that naïve beliefs regarding the source of internet-based medical knowledge had a negative relationship with the intention to use evidence-based online medical databases, that is, those health care professionals who believed that the internet is a good source of accurate and essential medical information were less likely to search for evidences from evidence-based online medical databases. On the contrary, sophisticated beliefs regarding justification for internet-based knowing in medicine had a positive relationship with intention to use evidence-based online medical databases. In other words, health care professionals who held beliefs that there is a need to justify web-based medical information by checking alternative sources were more likely to employ evidence-based online medical databases.

Consistently, it was indicated that source and justification play an essential role in web-based information searching to explore an ill-structured task. Specifically, the credibility of an electronic source and the criteria for the justification of the knowledge were frequently referred to when participants epistemically reflected during their web-based searching [27]. In particular, the source of knowledge is the most epistemic reflection during the web-based information search for a controversial topic. Obviously, the arousal of individuals' epistemic beliefs in source and justification is relatively important for improving their web-based searching behavior [24]. Accordingly, improving health care professionals' internet-specific beliefs regarding credibility of source and rule of justification may inspire them to access evidence from web-based medical databases.

Previous studies have shown that promoting the searching skills for retrieving and evaluating web-based medical information is essential for clinical health care professionals [8,11,14,50]. Besides training in internet search skills, there is a crucial need to elevate health care professionals' beliefs regarding internet-based knowledge and knowing in the medical domain. This study confirms the importance of improving epistemic beliefs to guide appropriate medical information behaviors [25,30]. Possessing advanced epistemic beliefs may stimulate information seekers to employ advanced strategies to evaluate the quality of web-based information [24]. Therefore, knowing health care professionals' internet-specific beliefs can help medical educators improve their web-based medical information–seeking behaviors and assist them in evaluating and retrieving the best evidence for clinical decision making.

Demographics and Intention to Use Evidence-Based Online Medical Databases

In addition to the influence of internet-specific epistemic beliefs on the intention to utilize the evidence-based medical databases, the demographics of health care professionals, including gender and academic degree, revealed a significantly influential role in the intention to search evidence-based online medical databases. Regarding the gender-related issue, gender differences in the use and perceptions of the internet remain a major concern when discussing web-based information-seeking behaviors [34,51]. Consistently, a gender difference was found in the intention to use web-based medical databases for evidence. Therefore, there is a need to investigate female health care professionals' needs for web-based information and to help them locate appropriate information in relation to evidence-based medicine. Finally, in accordance with other research studies, the results of this study showed that academic degree was an influential factor in the intention to use evidence-based online medical databases, that is, participants with a master's degree were more likely to search for evidence on medical databases. Perhaps their research training in their graduate courses gave them experience in searching web-based medical databases. Physicians' additional research degrees and research practice activities are associated with their evidence-based medicine competency [5]. Therefore, improving the knowledge of published research evidence and increasing participation in research training may be of potential benefit to conducting web-based searching for evidence-based medicine [6,52].

Limitations

This study has several major limitations that should be acknowledged. First of all, the generalizability of the study findings is limited. The participants surveyed were health care professionals in only 1 university-affiliated teaching hospital although it is a large-scale medical center. Second, the participants answered the paper-and-pencil questionnaire by themselves. Thus, self-reported bias may have occurred. Third, web-based medical databases are not the only reliable source of evidence-based medical information. Other alternative information sources of evidence-based medicine such as scientific, medical, health and nursing journals available in print cannot be underestimated and should be further studied. Fourth, the results of the structural model analysis showed a small R^2 value of 0.13. In addition to the predictors included in the structural model shown in Figure 1, in future studies, there may be some other variables, which can be treated as predictors, moderators, as well as mediators in health care professionals' intention to use evidence-based online medical databases. Finally, the RMSEA value of 0.071 revealed unsatisfactory model fit compared to the recommended threshold of 0.005 and below [53].

Conclusions

Evidence-based medicine plays a determinant role in health care quality. The internet has been regarded as an

evidence-based medical tool. Although the internet has become the most utilized resource for web-based medical information, health care professionals seldom access the validated evidence-based online medical databases. This study advances the knowledge on personal epistemic beliefs and their relationship with web-based information searching in clinical practice. Further, the results of this study provide suggestions for improving health care professionals' intention to utilize the evidence-based online medical databases.

Acknowledgments

This work was, in part, financially supported by the Institute for Research Excellence in Learning Sciences of National Taiwan Normal University from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan. In addition, the authors thank the research fund offered by the Ministry of Science and Technology (MOST), Taiwan, under grant numbers MOST 107-2511-H-002-010-MY3 and MOST 108-2511-H-003-038-MY3, and the fund provided by China Medical University, Taichung, under grant number CMU DMR-104-004. We also thank all those who were involved in this research, including the health care professionals who replied to the questionnaires and the research assistants who contributed to collecting the questionnaires. Finally, we deeply appreciate the reviewers who provided constructive and relevant comments on this study during the reviewing process.

Authors' Contributions

YLC contributed to the study design, developed the instruments, analyzed the research data, and drafted the main text of this paper. YCL provided opinions on the development of the research instrument. In addition, he assisted in delivering and collecting the questionnaires. CCT contributed to constructing the research model. He also gave opinions and interpreted the results of the statistical analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Statistical analyses of the participant data.
[DOCX File , 16 KB - jmir_v23i3e20030_app1.docx]

References

1. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 Jan 13;312(7023):71-72 [FREE Full text] [doi: 10.1136/bmj.312.7023.71] [Medline: 8555924]
2. Maggio LA, Kung JY. How are medical students trained to locate biomedical information to practice evidence-based medicine? A review of the 2007-2012 literature. J Med Libr Assoc 2014 Jul;102(3):184-191 [FREE Full text] [doi: 10.3163/1536-5050.102.3.008] [Medline: 25031559]
3. Schuers M, Griffon N, Kerdelhue G, Foubert Q, Mercier A, Darmoni SJ. Behavior and attitudes of residents and general practitioners in searching for health information: From intention to practice. Int J Med Inform 2016 May;89:9-14. [doi: 10.1016/j.ijmedinf.2016.02.003] [Medline: 26980354]
4. Al Omari M, Khader Y, Jadallah K, Dauod AS, Al-Shdifat AAK. Awareness, attitude and practice of evidence-based medicine among primary health care doctors in Jordan. J Eval Clin Pract 2009 Dec;15(6):1131-1136. [doi: 10.1111/j.1365-2753.2009.01223.x] [Medline: 20367716]
5. Buscaglia J, Nagula S, Yuan J, Bucobo JC, Kumar A, Forsmark CE, et al. The practice of evidence-based medicine (EBM) in gastroenterology: discrepancies between EBM familiarity and EBM competency. Therap Adv Gastroenterol 2011 Aug 06;4(5):283-294. [doi: 10.1177/1756283x11412240]
6. Heighes PT, Doig GS. Intensive care specialists' knowledge, attitudes, and professional use of published research evidence: a mail-out questionnaire survey of appropriate use of research evidence in clinical practice. J Crit Care 2014 Feb;29(1):116-122. [doi: 10.1016/j.jcrc.2013.10.014] [Medline: 24331946]
7. Mikalef P, Kourouthanassis PE, Pateli AG. Online information search behaviour of physicians. Health Info Libr J 2017 Mar;34(1):58-73 [FREE Full text] [doi: 10.1111/hir.12170] [Medline: 28168831]
8. Clarke MA, Belden JL, Koopman RJ, Steege LM, Moore JL, Canfield SM, et al. Information needs and information-seeking behaviour analysis of primary care physicians and nurses: a literature review. Health Info Libr J 2013 Sep;30(3):178-190 [FREE Full text] [doi: 10.1111/hir.12036] [Medline: 23981019]
9. Sarbaz M, Kimiafar K, Sheikhtaheri A, Taherzadeh Z, Eslami S. Nurses' Information Seeking Behavior for Clinical Practice: A Case Study in a Developing Country. Stud Health Technol Inform 2016;225:23-27. [Medline: 27332155]
10. Hornby K. The internet as an evidence-based medicine tool. BCMedJ. 2004. URL: https://bcmj.org/articles/internet-evidence-based-medicine-tool [accessed 2021-01-02]

11. Lialiou P, Mantas J. Online Information Seeking Behaviour by Nurses and Physicians: A Cross-Sectional Study. Stud Health Technol Inform 2016;225:33-37. [Medline: 27332157]
12. Younger P. Internet-based information-seeking behaviour amongst doctors and nurses: a short review of the literature. Health Info Libr J 2010 Mar;27(1):2-10 [FREE Full text] [doi: 10.1111/j.1471-1842.2010.00883.x] [Medline: 20402799]
13. Boruff JT, Harrison P. Assessment of knowledge and skills in information literacy instruction for rehabilitation sciences students: a scoping review. J Med Libr Assoc 2018 Jan;106(1):15-37 [FREE Full text] [doi: 10.5195/jmla.2018.227] [Medline: 29339931]
14. Hughes B, Joshi I, Lemonde H, Wareham J. Junior physician's use of Web 2.0 for information seeking and medical education: a qualitative study. Int J Med Inform 2009 Oct;78(10):645-655. [doi: 10.1016/j.ijmedinf.2009.04.008] [Medline: 19501017]
15. Maggio LA, Tannery NH, Chen HC, Cate OT, O'Brien B. Evidence-Based Medicine Training in Undergraduate Medical Education. Academic Medicine 2013;88(7):1022-1028. [doi: 10.1097/acm.0b013e3182951959]
16. Maggio LA, Aakre CA, Del Fiol G, Shellum J, Cook DA. Impact of Clinicians' Use of Electronic Knowledge Resources on Clinical and Learning Outcomes: Systematic Review and Meta-Analysis. J Med Internet Res 2019 Jul 25;21(7):e13315 [FREE Full text] [doi: 10.2196/13315] [Medline: 31359865]
17. Veness M, Rikard-Bell G, Ward J. Views of Australian and New Zealand radiation oncologists and registrars about evidence-based medicine and their access to Internet based sources of evidence. Australas Radiol 2003 Dec;47(4):409-415. [doi: 10.1046/j.1440-1673.2003.01211.x] [Medline: 14641194]
18. Liang JC, Tsai CC. The information commitments toward web information among medical students in Taiwan. Educational Technology & Society. 2009. URL: https://www.jstor.org/stable/10.2307/jeductechsoci.12.1.162 [accessed 2020-09-02]
19. Mokhtar IA, Majid S, Foo S, Zhang X, Theng Y, Chang Y, et al. Evidence-based practice and related information literacy skills of nurses in Singapore: an exploratory case study. Health Informatics J 2012 Mar;18(1):12-25 [FREE Full text] [doi: 10.1177/1460458211434753] [Medline: 22447874]
20. Weng Y, Kuo KN, Yang C, Lo H, Shih Y, Chen C, et al. Increasing utilization of Internet-based resources following efforts to promote evidence-based medicine: a national study in Taiwan. BMC Med Inform Decis Mak 2013 Jan 07;13:4 [FREE Full text] [doi: 10.1186/1472-6947-13-4] [Medline: 23289500]
21. Hofer BK, Pintrich PR. The Development of Epistemological Theories: Beliefs About Knowledge and Knowing and Their Relation to Learning. Review of Educational Research 2016 Jun 30;67(1):88-140. [doi: 10.3102/00346543067001088]
22. Chiu Y, Liang J, Hou C, Tsai C. Exploring the relationships between epistemic beliefs about medicine and approaches to learning medicine: a structural equation modeling analysis. BMC Med Educ 2016 Jul 18;16:181 [FREE Full text] [doi: 10.1186/s12909-016-0707-0] [Medline: 27431001]
23. Eastwood JL, Koppelman-White E, Mi M, Wasserman JA, Krug Iii EF, Joyce B. Epistemic cognition in medical education: a literature review. Int J Med Educ 2017 Jan 07;8:1-12 [FREE Full text] [doi: 10.5116/ijme.5849.bfce] [Medline: 28064257]
24. Ulyshen TZ, Koehler MJ, Gao F. Understanding the Connection Between Epistemic Beliefs and Internet Searching. Journal of Educational Computing Research 2015 Sep 30;53(3):345-383. [doi: 10.1177/0735633115599604]
25. Knight S, Rienties B, Littleton K, Mitsui M, Tempelaar D, Shah C. The relationship of (perceived) epistemic cognition to interaction with resources on the internet. Computers in Human Behavior 2017 Aug;73:507-518. [doi: 10.1016/j.chb.2017.04.014]
26. Mason L, Ariasi N, Boldrin A. Epistemic beliefs in action: Spontaneous reflections about knowledge and knowing during online information searching and their influence on learning. Learning and Instruction 2011 Feb;21(1):137-151. [doi: 10.1016/j.learninstruc.2010.01.001]
27. Mason L, Boldrin A, Ariasi N. Searching the Web to learn about a controversial topic: are students epistemically active? Instr Sci 2009 Jan 8;38(6):607-633. [doi: 10.1007/s11251-008-9089-y]
28. Hofer BK. Epistemological Understanding as a Metacognitive Process: Thinking Aloud During Online Searching. Educational Psychologist 2004 Mar;39(1):43-55. [doi: 10.1207/s15326985ep3901_5]
29. Chua AY, Banerjee S. To share or not to share: The role of epistemic belief in online health rumors. Int J Med Inform 2017 Dec;108:36-41. [doi: 10.1016/j.ijmedinf.2017.08.010] [Medline: 29132629]
30. Kienhues D, Stadtler M, Bromme R. Dealing with conflicting or consistent medical information on the web: When expert information breeds laypersons' doubts about experts. Learning and Instruction 2011 Apr;21(2):193-204. [doi: 10.1016/j.learninstruc.2010.02.004]
31. Kammerer Y, Amann DG, Gerjets P. When adults without university education search the Internet for health information: The roles of Internet-specific epistemic beliefs and a source evaluation intervention. Computers in Human Behavior 2015 Jul;48:297-309. [doi: 10.1016/j.chb.2015.01.045]
32. Bråten I, Strømsø HI, Samuelstuen MS. The Relationship between Internet-Specific Epistemological Beliefs and Learning within Internet Technologies. Journal of Educational Computing Research 2016 Jul 22;33(2):141-171. [doi: 10.2190/e763-x0ln-6nmf-cb86]
33. Chiu Y, Liang J, Tsai C. Internet-specific epistemic beliefs and self-regulated learning in online academic information searching. Metacognition Learning 2013 Jun 27;8(3):235-260. [doi: 10.1007/s11409-013-9103-x]
34. Chiu YL, Tsai CC, Liang JC. Testing measurement invariance and latent mean differences across gender groups in college students' Internet-specific epistemic beliefs. AJET 2015 Sep 13;31(4):486-499 [FREE Full text] [doi: 10.14742/ajet.1437]

35. Kammerer Y, Bråten I, Gerjets P, Strømsø H. The role of Internet-specific epistemic beliefs in laypersons' source evaluations and decisions during Web search on a medical issue. Computers in Human Behavior 2013 May;29(3):1193-1203 [FREE Full text] [doi: 10.1016/j.chb.2012.10.012]
36. Chiu Y, Liang J, Tsai C. Exploring the roles of education and Internet search experience in students' Internet-specific epistemic beliefs. Computers in Human Behavior 2016 Sep;62:286-291. [doi: 10.1016/j.chb.2016.03.091]
37. Hofer BK. Dimensionality and Disciplinary Differences in Personal Epistemology. Contemp Educ Psychol 2000 Oct;25(4):378-405. [doi: 10.1006/ceps.1999.1026] [Medline: 11001783]
38. Muis KR, Bendixen LD, Haerle FC. Domain-Generality and Domain-Specificity in Personal Epistemology Research: Philosophical and Empirical Reflections in the Development of a Theoretical Framework. Educ Psychol Rev 2006 Jul 29;18(1):3-54. [doi: 10.1007/s10648-006-9003-6]
39. Conley AM, Pintrich PR, Vekiri I, Harrison D. Changes in epistemological beliefs in elementary science students. Contemporary Educational Psychology 2004 Apr;29(2):186-204. [doi: 10.1016/j.cedpsych.2004.01.004]
40. Liang JC, Lee MH, Tsai CC. The Relations Between Scientific Epistemological Beliefs and Approaches to Learning Science Among Science-Major Undergraduates in Taiwan. Asia-Pac.Educ.Res 2010 Feb 25;19(1):43-59 [FREE Full text] [doi: 10.3860/taper.v19i1.1508]
41. Cullen RJ. In search of evidence: family practitioners' use of the Internet for clinical information. J Med Libr Assoc 2002 Oct;90(4):370-379 [FREE Full text] [Medline: 12398243]
42. Bråten I, Brandmo C, Kammerer Y. A Validation Study of the Internet-Specific Epistemic Justification Inventory With Norwegian Preservice Teachers. Journal of Educational Computing Research 2018 May 16;57(4):877-900. [doi: 10.1177/0735633118769438]
43. Hair JJF, Black WC, Babin BJ, Anderson RE, Tatham RL. Multivariate Data Analysis. 6th ed. NJ: Prentice Hall; 2006.
44. Chiu Y, Tsai C. The roles of social factor and internet self-efficacy in nurses' web-based continuing learning. Nurse Educ Today 2014 Mar;34(3):446-450. [doi: 10.1016/j.nedt.2013.04.013] [Medline: 23669599]
45. Hair Jr. JF, Matthews LM, Matthews RL, Sarstedt M. PLS-SEM or CB-SEM: updated guidelines on which method to use. IJMDA 2017;1(2):107. [doi: 10.1504/ijmda.2017.10008574]
46. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research 1981 Feb;18(1):39. [doi: 10.2307/3151312]
47. Anderson JC, Gerbing DW. Structural equation modeling in practice: A review and recommended two-step approach. Psychological Bulletin 1988 May;103(3):411-423. [doi: 10.1037/0033-2909.103.3.411]
48. Hair JJF, Hult GTM, Ringle CM, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). CA, USA: SAGE Publications Inc; 2016.
49. Kammerer Y, Gerjets P. Effects of search interface and Internet-specific epistemic beliefs on source evaluations during Web search for medical information: an eye-tracking study. Behaviour & Information Technology 2012 Jan;31(1):83-97. [doi: 10.1080/0144929x.2011.599040]
50. Scott SD, Gilmour J, Fielden J. Nursing students and Internet health information. Nurse Educ Today 2008 Nov;28(8):993-1001. [doi: 10.1016/j.nedt.2008.06.002] [Medline: 18700178]
51. Bidmon S, Terlutter R. Gender Differences in Searching for Health Information on the Internet and the Virtual Patient-Physician Relationship in Germany: Exploratory Results on How Men and Women Differ and Why. J Med Internet Res 2015 Jun 22;17(6):e156 [FREE Full text] [doi: 10.2196/jmir.4127] [Medline: 26099325]
52. Salbach N, Guilcher S, Jaglal S, Davis D. Factors influencing information seeking by physical therapists providing stroke management. Phys Ther 2009 Oct;89(10):1039-1050. [doi: 10.2522/ptj.20090081] [Medline: 19661160]
53. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]

Abbreviations

C-ISEQ: Chinese version of the internet-specific epistemic beliefs questionnaire
ISEBM: internet-specific epistemic beliefs in medicine
PLS-SEM: partial least squares-structural equation modeling
RMSEA: root-mean-square error of approximation
SEM: structural equation modeling

Edited by G Eysenbach; submitted 09.05.20; peer-reviewed by R Terlutter, A Vagelatos, S Bidmon; comments to author 08.08.20; revised version received 02.10.20; accepted 25.02.21; published 18.03.21.

Please cite as:
Chiu YL, Lee YC, Tsai CC
Internet-Specific Epistemic Beliefs in Medicine and Intention to Use Evidence-Based Online Medical Databases Among Health Care Professionals: Cross-sectional Survey
J Med Internet Res 2021;23(3):e20030
URL: https://www.jmir.org/2021/3/e20030
doi:10.2196/20030
PMID:33734092

Original Paper

Testing the Digital Health Literacy Instrument for Adolescents: Cognitive Interviews

Eunhee Park[1], PhD, RN; Misol Kwon[1], BS, RN

School of Nursing, University at Buffalo, Buffalo, NY, United States

Corresponding Author:
Eunhee Park, PhD, RN
School of Nursing
University at Buffalo
201 E Wende Hall, 3435 Main St
Buffalo, NY, 14214-3079
United States
Phone: 1 716 829 3701
Email: eunheepa@buffalo.edu

Abstract

Background: Despite the increasing number of youth seeking health information on the internet, few studies have been conducted to measure digital health literacy in this population. The digital health literacy instrument (DHLI) is defined as a scale that measures the ability to operate digital devices and read and write in web-based modes, and it assesses seven subconstructs: operational skills, navigation skills, information searching, evaluating reliability, determining relevance, adding self-generated content to a web-based app, and protecting privacy. Currently, there is no validation process of this instrument among adolescents.

Objective: This study aims to explore the usability and content validity of DHLI.

Methods: Upon the approval of institutional review board protocol, cognitive interviews were conducted. A total of 34 adolescents aged 10-18 years (n=17, 50% female) participated in individual cognitive interviews. Two rounds of concurrent cognitive interviews were conducted to assess the content validity of DHLI using the *thinking aloud* method and probing questions.

Results: Clarity related to unclear wording, undefined technical terms, vague terms, and difficult vocabularies was a major issue identified. Problems related to potentially inappropriate assumptions were also identified. In addition, concerns related to recall bias and socially sensitive phenomena were raised. No issues regarding response options or instrument instructions were noted.

Conclusions: The initial round of interviews provided a potential resolution to the problems identified with comprehension and communication, whereas the second round prompted improvement in content validity. Dual rounds of cognitive interviews provided substantial insights into survey interpretation when introduced to US adolescents. This study examined the validity of the DHLI and suggests revision points for assessing adolescent digital health literacy.

(J Med Internet Res 2021;23(3):e17856) doi:10.2196/17856

KEYWORDS

adolescent; digital health literacy; ehealth literacy; cognitive interview

Introduction

Background

According to the recent report by the International Telecommunications Union, the majority of the world population (93%) resides within reach of mobile broadband (or internet) service [1]. Within this population, it is estimated that 53.6% people–approximately 4.1 billion people—are connected to the internet and have internet access at home [1]. In recent years, digital devices such as mobile phones, tablets, and portable computers have become essential tools in health care. Digital devices have become a medium for improving and facilitating health and well-being, and providing access to health care services for individuals and across populations [2]. The public uses these devices to obtain health-related information on the web and communicate with people using email and social media platforms [3]. In addition, web- and app–based tools aid digital devices in becoming increasingly resourceful in delivering health education and interventions, and empowering users [4]. Given the growth of technology, digital technology-based interventions provide opportunities to improve

access to care and create positive health outcomes by enhancing patient engagement and self-management skills using tailored programs and interactive features [5,6].

In particular, youth spend significant time using digital devices such as mobile phones and portable tablets, not only in their daily lives but also for obtaining health-related information and participating in social media for peer support and self-care [3,7,8]. Adolescents in our era are *digital natives* who are comfortable with and attracted to technology and the use of digital devices. It has been reported that teens have a high demand for and openness to the use of digital devices for their health and self-care behavior [8-10]. However, despite general knowledge, they still need to be educated about the specific use of these devices, especially in terms of understanding and deciphering the surplus of content on the web. Thus, it is important to consider digital health literacy and provide proper education and interventions to enhance skills in maneuvering and making decisions based on the information found on the web; this is expected to improve the ability to appropriately use digital devices in the context of health, as they are closely related to mobile health usability and may influence the outcomes of digital-based health care intervention and education. Subsequently, as a high percentage of information is provided on the web for youth—and they are inclined to use digital modes of communication—, such efforts would influence young people's health behaviors and create positive outcomes [11-15].

To capture how youth interpret digital health information and use multiple digital devices for health, the notion of digital health literacy is most frequently defined as the ability to find and understand health-related information on the web and to write and post interactive features on the internet [16]. Digital health literacy is similar to the concept of *eHealth literacy*, which refers to the ability to read and write in web-based modes requiring multiple components from information, science, health, media, and computer literacy [17-20]. Unlike eHealth literacy, which merely focuses on the ability to read and write information on the web based on *health 1.0 skills*, digital health literacy expands these concepts by including the skills needed to write and post health-related messages on the web based on *health 2.0 skills* [17]. The terms *health 1.0* and *health 2.0* originated from the concept of *web 2.0* in the domain of health, and *health 2.0* indicates advanced technology involving patient empowerment and involvement, sharing information, and social networking [21]. This is important given the development of social media features and web-based discussion threads, which not only require the ability to search for health-related information, understand the information, and apply it appropriately but also to write and post their own information on the web [22].

Given the importance of understanding and applying health information on the web, health researchers have made efforts to develop instruments to assess digital health literacy. However, most of these instruments have only focused on specific aspects of digital health literacy, particularly with regard to health 1.0 skills. For instance, Rapid Estimate of Adolescent Literacy in Medicine-Teen [23] assesses the literacy level to correctly pronounce health-specific terms but does not include an assessment of the understanding of those terms. Newest Vital Sign focuses on understanding numeric information in the context of health [24]. The eHealth Literacy Scale assesses young people's perceived health literacy regarding finding, assessing, and understanding electronic health information but does not assess the ability to write messages or posts on the web or manage digital devices [17,25]. Digital Health Literacy Assessment Tool assesses adolescents' ability to find and assess health-related information on the web [26,27].

Objectives

Considering the limitations of the existing instruments, the digital health literacy instrument (DHLI) [16] assesses an expanded concept of digital health literacy, assessing seven skills: (1) operation skills, (2) navigation skills, (3) information searching, (4) evaluating reliability, (5) determining relevance, (6) adding self-generated content, and (7) protecting privacy, which were identified as necessary skills needed for health-related internet use [16,28]. This instrument, designed for adult users, offers many components relevant to the assessment of digital health literacy based on health 2.0 skills. We will use a DHLI, which is an instrument developed to assess digital health literacy, as a model for our research focused on the adolescent population. The scale comprises 21 self-reported items for adults. It was originally developed in Dutch and translated into English, and the developers conducted cognitive interviews to ensure its validity in adults [29,30]. Thus, its reliability and validity are supported for the Dutch adult population [16]. The internal consistency of the total scale (Cronbach α=.87) and the subscales (Cronbach α=.70-.89) was supported, but the value for the privacy protection subscale (Cronbach α=.57) was low. A test-retest analysis was conducted, and intraclass correlation coefficients for the total scale and subscales ranged from .49 to .81 [16]. Confirmatory factor analysis supported the 7 constructs. Correlations with significant associations were reported for age, education, internet use, health-related internet use, self-reported health status, health literacy, and eHealth to support validity of the instrument [16].

Although digital health literacy is an important concept, there are limited instruments available for adolescents. The lack of existing instruments that effectively assess adolescents' digital health literacy has led to a limited understanding of young people's comprehension skills and the ability to use the internet and digital devices for health-related purposes [27]. Unfortunately, this reduces the chances of successful results of prevalent adolescent health education and health promotion programs with digital devices. On the basis of these concerns, we aim to explore usability and assess the validity of DHLI using cognitive interviews among US adolescents (aged 10-18 years). Cognitive interviews are an essential step in ensuring the validity of the questionnaire development process. Cognitive interviews provide a chance to find sources of errors in survey research and allow for exploration of the understanding of the questionnaire among the target population [28]. Thus, the findings of this study will contribute to modifying DHLI and tailoring the instruments targeting adolescents in the United States.

Methods

Research Design

After approval from the affiliated university, cognitive interviews were conducted. Parent consent forms and minor assent forms were obtained. Concurrent cognitive techniques have been used [31]. Two interviewers were trained for the cognitive interviews with adolescents. The first author, who had experience conducting cognitive interviews related to previous works, provided the essential literature on cognitive interviews, including books, example articles, and an interview guide, to the other research team members (2 interviewers). The research team members read the selected books and articles, after which the first author guided the interview process step by step. We used the training procedure suggested by Willis [32]. The research team members met regularly (>3 times) during the 2 weeks of training sessions until the interviewers felt comfortable with and confident about the procedures and were able to conduct the interview on their own. The teach-back method, in which a trained interviewer teaches and prompts the training interviewer as guided, was used, allowing for a standardized process. After each interview, we listened to each other's interview audio recordings and provided feedback and suggestions for improvement when necessary. In addition, we met on a weekly basis and discussed elements of the interview process, such as clarifying procedures, resolving any obstacles, and discussing probing questions. The interview process consisted of one-on-one individual interviews, and the audio was recorded.

Participants

A total of 34 adolescents completed the interviews. The sample was composed of 50% (n=17) females and 50% (n=17) males aged between 10 and 18 years. The ethnic background included White Americans (n=15, 44%), African Americans (n=11, 32%), Latino (n=3, 9%), and Asian Americans (n=5, 15%). The mean age of the participants was 13.47 (SD 2.39) years. Table 1 provides detailed demographic information of the respondents. Participants were recruited via flyers posted during summer at the youth community centers and universities where adolescents often visit for summer camp programs.

Table 1. Demographics of the recruited adolescents (N=34).

Variables	Participants, n (%)
Gender, n (%)	
Female	17 (50)
Male	17 (50)
Race, n (%)	
White American	15 (44)
Black or African American	11 (32)
Latino or Hispanic	3 (9)
Asian American	5 (15)
Age (years), mean (SD)	13.47 (2.39)
School achievement, n (%)	
Much better than average	3 (9)
Better than average	16 (47)
Average	14 (41)
Much worse than average	1 (3)
Literacy level, n (%)	
Much better than average	8 (24)
Better than average	13 (38)
Average	13 (38)
Health insurance type, n (%)	
Private	4 (12)
Government-sponsored	7 (21)
Does not know	23 (68)
Free lunch at school, n (%)	
Yes	21 (62)
No	13 (38)

We used a quota sampling approach for sample selection [32,33]. We aimed to recruit a similar proportion of male and female participants (50% each). Recruitment also accounted for the racial and ethnic composition of the US population. For example, according to the census data from 2018, the US population was 61% White, 13% Black or African American, 17% Latino or Hispanic (17%), 5% Asian, 1% American Indian or Alaska Native, and 1% Native Hawaiian or other Pacific Islander. We attempted to recruit participants corresponding to the number of participants for each racial and ethnic group. Among our participants, there was a slightly higher percentage of Black or African Americans and Asian Americans and a lower percentage of White Americans and Latino or Hispanics than we had intended. We could not recruit participants from American Indian, Alaska Native, Native Hawaiian, or other Pacific Islander groups. In addition, we recruited from all age groups, resulting in at least 3% of all participants in each age group from 10 to 18 years.

Data Collection

Each face-to-face interview was conducted in English and lasted for 30 minutes to 1 hour. Before each interview, the researcher explained the purpose and process of the interview to the participants. In the first round, 22 interviews were conducted. On the basis of the findings from the first round of interviews, the research team analyzed the data and made a suggested revision to the instruments. In the second round, we conducted 12 interviews with the original and revised items. According to Willis [33], 12-15 is a good target number of interviews for each round. However, we felt the need to include more than 15 interviews to ensure a diverse population of adolescents, especially those of various ages [33]. Therefore, we had 22 participants. We stopped recruiting participants when saturated themes emerged, which indicated that no more new issues or themes arose during the interviews.

The participants read each item and were encouraged to think aloud with regard to each item. The researcher then asked probing questions based on the previous literature [32,34]. All interview files were transcribed verbatim.

Data Analysis

The researchers listened to the audio recordings, and the transcribed data were coded based on the Miles and Huberman approach [35]. Using the Miles and Huberman approach, we followed a deductive approach to analyze the interview data. We formulated a clear research question and used the theoretical framework to guide the study. This approach was chosen because of the reasoning that the nature of our study would benefit from this analytic approach with the existing framework and facilitate the assessment of the instrument's content validity in a more structured way. We initially conducted descriptive coding by labeling each unit of meaning related to specific issues raised for each questionnaire item. Guided by the Willis [32] framework, we established the following main categories: reading, instruction, clarity, assumptions, knowledge or memory, sensitivity or bias, response categories, and other problems [32-34]. The following categories and subdomains of the Willis [32] framework (Table 2) were used to ensure the instrument's content validity when organizing the codes. In addition, the framework provides clear categories with subdomains, which are important for ensuring content validity in a survey development process based on user feedback from existing questionnaires. The authors checked for any emergent codes in the data, but none were found.

Table 2. Results of cognitive interviews.

Category and subdomains	Original item	Revised items
Clarity		
Wording		
	5a. Do you find it difficult to judge who can read along?	Do you find it difficult to *know who will read the message*?[a]
Technical terms		
	1c. Do you use the buttons or links and hyperlinks on websites?	Do you use the buttons or links on websites?
	4. When typing a message (eg, to your doctor, on a forum or on social media such as Facebook or Twitter) how easy or difficult is it for you to...	When typing a message *online* (eg, to your doctor; on a *website*; *blog*; or on social media such as Facebook, Twitter, *Snapchat*, or *Instagram*) how easy or difficult is it for you to...
	5. When you post a message on a public forum or social media, how often...	When you *write* a message on a *website*, *blog*, or social media, how often...
Vague		
	2e. Decide whether the information is written with commercial interests (eg, by people trying to sell a product).	Decide whether the information is written *for advertisement* (eg, by people trying to sell a product).
	2g. Decide if the information you found is applicable to you.	Decide if the information you *find relates* to you (eg, *school homework*, *exercise*, and *eating habits*).
	2h. Apply the information you found in your daily life.	Apply the information you *find* in your daily life (eg, *school homework*, *exercise*, and *eating habits*).
	5b. Do you (intentionally or unintentionally) share your own private information (eg, name or address)?	Do you share your private information (eg, name or address, *location*, and *school information*)?
	5c. Do you (intentionally or unintentionally) share some else's private information?	Do you share *someone* else's private information (eg, name or address, *location*, and *school information*)?
Difficult vocabulary		
	2b. Use the proper words or search query to find the information you are looking for.	Use the *keywords* or search *term* to find the information you are looking for.
	2d. Decide whether the information is reliable or not.	Decide whether the information is *trustworthy* or not.
	2g. Decide if the information you found is applicable to you.	Decide if the information you found *relates* to you (eg, *school homework*, *exercise*, and *eating habits*).
	4a. Clearly formulate your question or health-related worry.	Clearly *write* your question or health-related worry.
Assumption		
Inappropriate assumptions		
	1a. Use the keyboard of a computer (eg, to type words).	Use the keyboard of a computer, or a *tablet*, or a *phone* (eg, to type words).
	1b. Use the mouse (eg, to put the cursor in the right field or to click).	Use the mouse or a *touchpad* (eg, to put the cursor in the right field or to click).
	4. When typing a message (eg, to you doctor, on a forum or on social media such as Facebook or Twitter) how easy or difficult is it for you to...	When typing a message *online* (eg, to your doctor; on a *website*; *blog*; or on social media such as Facebook, Twitter, *Snapchat*, or *Instagram*) how easy or difficult is it for you to...
	5. When you post a message on a public forum or social media, how often...	When you *write* a message on a *website*, *blog*, or social media, how often...
Knowledge		
Recall	2h. Apply the information you found in your daily life.	Apply the information you *find* in your daily life (eg, *school homework*, *exercise*, and *eating habits*).
Sensitivity/bias		
Socially acceptable		
	5b. Do you (intentionally or unintentionally) share your own private information (eg, name or address)?	Do you share your private information (eg, name or address, *location*, and *school information*)?
	5c. Do you (intentionally or unintentionally) share some else's private information?	Do you share *someone* else's private information (eg, name or address, *location*, and *school information*)?

[a]Please see the text in italics for the revised phrases.

Results

Overview of the Results

We conducted 34 cognitive interviews across 2 rounds to evaluate clarity, assumptions, knowledge, and sensitivity or bias in the items. Subthemes that emerged from the 4 categories included wording, technical terms, vagueness, difficult vocabulary, inappropriate assumptions, recall, and being socially acceptable. The initial round of interviews provided a potential resolution to comprehension and communication problems. We determined the need to propose a modification when participants repeatedly brought up similar issues during the initial round. Subsequently, the research team modified the items using collaborative clinical judgment. We had no specific quantitative cut-off points (eg, the minimum percentage of participants who raised each issue) because no such numbers are suggested in the literature, given that most cognitive interviews involve a small number of participants. When the minimum percentage of the participants who raised the same issue during the initial round (n=22) for the modified items was calculated, it was 18% (n≥4). During the modification phase, the researchers reflected on the participants' responses and suggestions. For instance, modifications were made by adding examples that were reflective and inclusive of current technological devices (eg, mouse vs mouse and touchpad). Most issues that indicated the need for improvement in the areas of clarity, assumptions, and knowledge emerged during the first round, whereas the second round provided a more confirmatory process, prompting improvement in content validity. Examples of changes for each category are presented in Table 2.

Clarity

Wording

Clarity of the items requires wording that is most appropriate to best explain a question being asked. We defined problematic wording as that which is seemingly lengthy, awkward, ungrammatical, or contains complicated syntax for the respondents. For instance, half of the respondents in round 1 (11/22, 50%) expressed awkwardness in the word *to judge* in the item 5a: "When you post a message on a public forum or social media, how often do you find it difficult to judge who can read along?"

When asked to interpret in their own words, many found the item difficult to comprehend and stated that it did not make sense even after reading it several times. A respondent stated:

> *I think I'm just getting confused at do you find it difficult to judge the community along – who's your – what is it asking?*

In addition, participants expressed confusion by the phrase *who can read along*. In round 2, we provided 3 options: wordings in the item that remained the same and later ones that replaced *to judge* with *to know* and *to decide*. *To know* was better received by the majority (over 8/12, 65%) of the respondents.

Technical Terms

The original items included undefined, unclear, or complex terms that were either redundant or needed more examples that best catered to the adolescent population. Over half of the respondents (13/22, 59%) struggled to define the term *hyperlinks* (item 1c); hence, it was removed. Likewise, several participants (7/22, 32%) had varied definitions of *forum* (item 4), ranging from a *form/document* to a *place where professionals gather*. Hence, the word *forum* was replaced with *websiteorblog*. Moreover, the participants (7/22, 32%) found Facebook and Twitter as examples of social media networking websites in the original survey (item 4) to lack relatability and suggested the inclusion of Snapchat and Instagram.

Vague

Vagueness can be a linguistic and nonlinguistic mechanism that enables readers to interpret questions based on their own understanding. This leads to interpreting the question in multiple ways, deciding what is to be included or excluded, or being selective based on their own definition, all of which prohibits a clear understanding. In round 1, many respondents (14/22, 64%) verbalized that *commercial interests* would be hard to pinpoint, especially if no examples were to be provided (item [2e]):

> *When you search the internet for information on health, how easy or difficult is it for you to decide whether the information is written with commercial interests?* [eg, by people trying to sell a product]

Respondents communicated that either replacing *commercial interest* with *advertisement* or shortening the question to "...do the commercials influence you during web-surfing" would yield more accurate responses.

Similarly, respondents (4/22, 18%) suggested the inclusion of examples that were more specific to children and adolescents to yield better responses. Hence, examples such as *school homework, exercise, eating habits* was added to item 2g and 2h, and *location, school information* were added to item 5b, and 5c, respectively. In round 2, respondents were probed on their confidence in providing an accurate understanding of these items, and they showed a better understanding with these examples. A respondent provided feedback by stating:

> *I like how you put the examples. I think that really helps. Because someone else can completely understand what that means. I think you just feel like ah like this is all about my homework. All this helped me to choose what food to eat. So I think it was really important for the examples. That really helps.*

Difficult Vocabulary

Text comprehension can be influenced by the quantity and location of difficult vocabulary in a given statement. Respondents (10/22, 45%) felt that words such as *applicable* (item 2g) and *formulate* (item 4a) could be replaced with easier words. This was especially evident with younger respondents (aged 10-12 years) who did not understand what the words meant but vaguely tried to guess in their own ways. Suggestions

were taken into consideration for round 2. With changes in round 2, all participants (n=12) expressed no difficulty in understanding. In addition, the *proper words* described in item 2b were confusing to the respondents (4/22, 18%). This was taken into consideration in round 2, and options were provided that replaced *proper words* with *key words*, *appropriate (key) words*, or *correct (key) words*. All options were provided, and most of the respondents (9/12, 75%) found *key words* to be an easier option for conveying the meaning embedded in the item.

In round 1 of cognitive interviews, several participants (11/22, 50%) indicated that the word *reliable* is understandable, but replacing it with *trustworthy* would be easier to understand, as it is more commonly used by their age group. In addition, more than half of the respondents (12/22, 55%) reported difficulty with the vocabulary—the *query*. Many had difficulty rephrasing the word or the item, as also described by one of the older respondents:

> *I understand and I know what a search query is. It just throws me off a little bit because it's not a very... it's not commonly used jargon for most people that I interact with. So that threw me off a little bit when you first read it to me.* [Aged 18 years, female]

The *search query* was then replaced with *term* for round 2, and all respondents (n=12) were able to comprehend the item with ease.

Assumptions: Inappropriate Assumptions

Survey questions require wording or phrases that do not make assumptions or draw conclusions about the respondent or his or her circumstances. For instance, the original survey of item 1a asks, "How easy or difficult is it for you to use the keyboard of a computer (eg, to type words)?" Early in round 1 of cognitive interviews, 18% (4/22) of the respondents suggested adding a computer, tablet, or phone, as teens nowadays increasingly use a phone or tablet more than a computer. When asked to apply the question to the individual, another respondent replied:

> *I use my fingers to type on the, on my phone, and my tablet, and stuff.*

Hence, *touch pad* was added to the original item 1b, "How easy or difficult is it for you to use the mouse (eg, to put the cursor in the right field or to click)?", and was rephrased to *mouse or touch pad* in round 2. In addition, contrary to the general belief, many younger respondents in the study stated that although they possess a social media account, they do not use them regularly.

Knowledge: Recall

The ability to answer survey questions appropriately and precisely requires respondents to be able to recall instances in their lives to answer the question. For instance, 18% (4/22) of the respondents found the original survey item 2h, "When you search the internet for information on health, how easy or difficult is it for you to apply the information you found in your daily life?" to be difficult to answer, as a respondent quotes:

> *I feel like it's asking a lot because we might not encounter information one day but if you do it a week from now, you still would remember it so you could use it later, so it's not really, you can't really answer the question because you don't know until you run across the information, like that's really a good fact. You remember it and you use it later on.*

This item was revised to read "apply the information you find in your daily life" by changing *found* to *find*. This emphasized perceived real-life applicability, as opposed to the actual application, to reduce recall bias and enhance the item's validity. However, we were unable to gather feedback on this change during the second round because we came up with this idea after all interviews.

In addition to having difficulty defining what *daily* entails, respondents felt that adding examples that pertain to the question could help them recall better and more quickly. In round 2, we added "(eg, school homework, exercise, eating habits)" at the end of item 2h.

Sensitivity or Bias: Socially Acceptable

We probed the respondents in an effort to understand whether the survey questions were sensitive in nature or wordings implicated bias. In answering the question about the frequency of sharing private information (eg, name or address) unintentionally and intentionally, most of the respondents confidently responded with *never*. However, when probed regarding the information and posts by their friends or themselves in their social media networking websites or apps, several respondents (4/22, 18%) recanted their response from *never* to *sometimes*, which may indicate potential bias and sensitivity in this question. A respondent stated:

> *You don't know everybody on social media but majority of people you do know. So like if you say going to the mall then maybe your friend might comment, "I want to come along." But you don't know who is seeing it. So it is kind of private but we don't think of it like that. But like if you really sit down and think about it then you can understand. But if we are typing, we are just like putting it up for our friends to see and not for everybody - but it is social media so everybody is going to see it. Like their name, their age, like what school they go to. But I've been seeing a lot of like – because I only know the 8th grades and 9ths so everybody go to different high schools and stuff but so they'll be like oh such and such guys have been to this school or this school so yeah.*

The privacy issue is pertinent when discussing digital health literacy, as youth often forget how much private information about themselves or others they expose or disclose when posting a public message. We suggested deleting *intentionally or unintentionally* from the original item to reduce potential sensitivity or bias related to differences between intentional and unintentional incidences of privacy breeches and clarify the item.

Discussion

Principal Findings

This study conducted cognitive interviews on the DHLI originally developed for adults in Dutch to tailor it to adolescents

in the United States. Cognitive interviews were conducted to assess content validity and identified issues based on four categories: clarity, assumptions, knowledge or memory, and sensitivity or bias. On the basis of the following issues, we made suggested revisions to the original items and received feedback in the second round of interviews.

Given that this instrument was originally developed for adults in Dutch and translated into English, clarity was the main issue found. We identified difficult vocabularies, technical terms, and wordings and vague expressions, which prevented a clearer understanding of each item. This finding suggests that back-translation may be necessary to ensure cultural validity in translating instruments.

Given that the digital world is rapidly evolving, some terms are likely to be outdated and may not reflect the current trends. For example, the original items include *mouse* and *computer* and did not include *touch pad* or *tablet*, which suggests that this type of instrument needs to reflect the rapidly changing trend of technology development. This has emerged concerning social-media-related questions. Adolescents did not use Facebook as much as other social media networking websites, such as Snapchat or Instagram. As trending websites and apps tend to evolve rapidly, instruments such as DHLI need to be able to reflect fast-changing trends in a timely manner.

Moreover, social media use or web searches may not be considered as frequent activities among some adolescents, particularly the younger population, which could lead to a potential threat to validity. Thus, we may not assume that all adolescents use social media or the internet for health-related purposes. It is also important to provide specific examples that reflect the context or activities familiar to adolescents when developing items, which will make the items more relevant to adolescents. For example, homework, school information, and eating habits would be examples that could be relevant to adolescents and could be added to the items.

Limitation and Suggestions for Future Research

This study has some limitations. Although we included 34 participants and aimed at diversity (age, sex, and race or ethnicity) in recruitment, the study included a slightly higher percentage of minority groups (African American, Hispanic, and Asian Americans). In addition, all participants were recruited from the same region (an eastern state in the United States), which may have decreased the generalizability of the findings. Future studies may also need to include participants who may not go to school or for whom English may not be their first language.

In addition, we admit that the developmental span varies even among adolescents across age. Thus, further differentiation may need to be considered—altering the questions to younger and older adolescents to consider their developmental stage and contextual factors and to develop more valid instruments.

Although the DHLI comprehensively assesses the skills necessary for digital health literacy, we only used the part of the scale that is based on the self-report. We were not able to tailor and test the 7 performance-based items in this cognitive interview because those items were provided by Dutch websites. This part of the scale is important to provide insightful understanding of digital health literacy by allowing the assessment of participants' performance level. It will be difficult to use these items as they are, but it will be important to develop this type of performance-based test for adolescents in the United States.

Another limitation is that this study focused on exploring the instrument's content validity based on cognitive interviews with adolescent users. Thus, based on the feedback we obtained from adolescents, providing quantitative data on the revised instrument's reliability and construct validity for adolescents is necessary for future studies. In addition, the suggested revisions are based only on adolescents' feedback on the translated scale. We made considerable efforts to preserve the original items as much as possible because of the validity and reliability established in previous studies. Nonetheless, further modifications may be necessary to enhance the scale's reliability and validity. For example, the original version of question 5 does not specifically contain the word *health*. Although the item presumes that the respondents know the questionnaire's purpose and that the item asks about health-related contexts, such an assumption may threaten the scale's validity. Further improvements might increase the scale's validity, as might exploring the revised instrument with experts.

Implications for Practice

In practice, health care professionals who develop interventions using the internet or digital devices may be able to assess digital health literacy levels among adolescents. Thus, the initial level of digital health literacy can be considered in the delivery of such interventions or education, and the program can be tailored based on the level of literacy [36]. This is a necessary step considering that digital device-based interventions have great advantages and have been beneficial for improving health in various contexts and have been particularly helpful given the nature of attractiveness to adolescents.

Conclusions

This study fills the important gap in research by exploring the validation of a DHLI for adolescents. Digital health literacy is an important skill that needs to be assessed, enhanced, and considered in the digital era. On the basis of cognitive interviews, the validity of the instrument assessing comprehensive digital health literacy skills was tested, and the items for adolescents with suggested revisions were provided (Multimedia Appendix 1). This will allow for the provision of tailored health education and promotion programs based on individual digital health literacy levels, and personalized effort will increase the chances for better health outcomes for this population.

Acknowledgments

The authors would like to thank Mary Rose Gaughan for her assistance during the data collection phase of the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Revised items.
[DOCX File , 17 KB - jmir_v23i3e17856_app1.docx]

References

1. International TU. Measuring digital development: Facts and figures. Geneva: ITU Publication; 2019.
2. Kostkova P. Grand challenges in digital health. Front Public Health 2015 May;3:134 [FREE Full text] [doi: 10.3389/fpubh.2015.00134] [Medline: 26000272]
3. Lenhart A. Pew Research Center. 2015. URL: http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/ [accessed 2021-03-01]
4. Older and newer media: patterns of use and effects on adolescents? health and well-being. J Res Adolesc 2011;21(1):95-113. [doi: 10.1111/j.1532-7795.2010.00717.x]
5. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013 Jan;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
6. Kim H, Park E, Lee S, Kim M, Park EJ, Hong S. Self-Management of Chronic Diseases Among Older Korean Adults: An mHealth Training, Protocol, and Feasibility Study. JMIR Mhealth Uhealth 2018 Jun 29;6(6):e147. [doi: 10.2196/mhealth.9988]
7. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. Soc Sci Med 2005 Apr;60(7):1467-1478. [doi: 10.1016/j.socscimed.2004.08.010] [Medline: 15652680]
8. Park E, Kwon M. Health-related internet use by children and adolescents: systematic review. J Med Internet Res 2018 Apr 03;20(4):e120. [doi: 10.2196/jmir.7731]
9. Prensky M. Digital native, digital immigrants. On the Horizon 2001;9(5):1-6.
10. Wong CA, Merchant RM, Moreno MA. Using social media to engage adolescents and young adults with their health. Healthc (Amst) 2014 Dec;2(4):220-224 [FREE Full text] [doi: 10.1016/j.hjdsi.2014.10.005] [Medline: 25984444]
11. Cha E, Kim KH, Lerner HM, Dawkins CR, Bello MK, Umpierrez G, et al. Health literacy, self-efficacy, food label use, and diet in young adults. Am J Hlth Behav 2014 May 1;38(3):331-339. [doi: 10.5993/ajhb.38.3.2]
12. Chari R, Warsh J, Ketterer T, Hossain J, Sharif I. Association between health literacy and child and adolescent obesity. Patient Educ Couns 2014 Jan;94(1):61-66. [doi: 10.1016/j.pec.2013.09.006] [Medline: 24120396]
13. 20 Health Literacy. National Network of Libraries of Medicine (NN/LM). URL: https://nnlm.gov/outreach/consumer/hlthlit.html [accessed 2021-02-26]
14. DeWalt DA, Hink A. Health literacy and child health outcomes: a systematic review of the literature. Pediatrics 2009 Oct 27;124(Supplement 3):S265-S274. [doi: 10.1542/peds.2009-1162b]
15. Park E, Drake E. Systematic review: internet-based program for youth smoking prevention and cessation. Journal of Nursing Scholarship 2014 Aug 18;47(1):43-50. [doi: 10.1111/jnu.12104]
16. van DV, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
17. Norman CD, Skinner HA. Ehealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
18. Ghaddar S, Valerio M, Garcia C, Hansen L. Adolescent health literacy: the importance of credible sources for online health information. J Sch Health 2012 Jan;82(1):28-36. [doi: 10.1111/j.1746-1561.2011.00664.x] [Medline: 22142172]
19. Higgins J, Begoray D. Exploring the Borderlands between Media and Health: Conceptualizing“ Critical Media Health Literacy”. Journal of Media Literacy Education 2012;4(2):148.
20. Subramaniam M, St Jean B, Taylor NG, Kodama C, Follman R, Casciotti D. Bit by bit: using design-based research to improve the health literacy of adolescents. JMIR Res Protoc 2015 May 29;4(2):e62 [FREE Full text] [doi: 10.2196/resprot.4058] [Medline: 26025101]
21. Van De Belt TH, Engelen LJ, Berben SA, Schoonhoven L. Definition of Health 2.0 and Medicine 2.0: a systematic review. J Med Internet Res 2010 Jun 11;12(2):e18 [FREE Full text] [doi: 10.2196/jmir.1350] [Medline: 20542857]
22. Kwon M, Park E. Perceptions and sentiments about electronic cigarette on the social media: a systematic review. JMIR Public Health and Surveillance. JMIR Publications 2020;6(1):e13673. [doi: 10.2196/preprints.13673]

23. Davis TC, Wolf MS, Arnold CL, Byrd RS, Long SW, Springer T, et al. Development and validation of the Rapid Estimate of Adolescent Literacy in Medicine (REALM-Teen): a tool to screen adolescents for below-grade reading in health care settings. Pediatrics 2006 Dec 01;118(6):e1707-e1714. [doi: 10.1542/peds.2006-1139] [Medline: 17142495]
24. Warsh J, Chari R, Badaczewski A, Hossain J, Sharif I. Can the Newest Vital Sign be used to assess health literacy in children and adolescents? Clin Pediatr (Phila) 2014 Feb 24;53(2):141-144. [doi: 10.1177/0009922813504025] [Medline: 24065737]
25. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
26. Subramaniam M, Taylor N. As simple as that?: Tween credibility assessment in a complex online world. J Doc 2015;71(3):550-571. [doi: 10.1108/jd-03-2014-0049]
27. St. Jean B, Greene Taylor N, Kodama C, Subramaniam M. Assessing the digital health literacy skills of tween participants in a school-library-based after-school program. Journal of Consumer Health on the Internet 2017 Mar 10;21(1):40-61. [doi: 10.1080/15398285.2017.1279894]
28. van der Vaart R, Drossaert CH, de Heus M, Taal E, van de Laar MA. Measuring actual eHealth literacy among patients with rheumatic diseases: a qualitative analysis of problems encountered using Health 1.0 and Health 2.0 applications. J Med Internet Res 2013 Feb 11;15(2):e27 [FREE Full text] [doi: 10.2196/jmir.2428] [Medline: 23399720]
29. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
30. Hak T, Van DV, Jansen H. The Three-Step Test-Interview (TSTI): an observation-based method for pretesting self-completion questionnaires. Survey Research Methods 2008;2(3):a.
31. Rothgeb J, Willis G, Forsyth B. Questionnaire pretesting methods: do different techniques and different organizations produce similar results? Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique 2016 Jul 22;96(1):5-31. [doi: 10.1177/0759106307096001003]
32. Willis GB. Cognitive Interviewing: A “How To” Guide: A Tool for Improving Questionnaire Design: SAGE Publications, Inc; 2005.
33. Willis GB. Cognitive aspects of survey methodology. In: Lavrakas P, editor. Encyclopedia of Survey Research Methods. Vol 2. Thousand Oaks, CA: Sage Publications; 2009:103-106.
34. Willis GB. Developing standard cognitive probes. In: Cognitive Interviewing. Thousand Oaks, CA: SAGE Publications, Inc; 2011:66-86.
35. Miles M, Huberman A. Qualitative Data Analysis. Thousand Oaks, CA: Sage; 1994.
36. Cushing CC, Steele RG. A meta-analytic review of eHealth interventions for pediatric health promoting and maintaining behaviors. J Pediatr Psychol 2010 Oct;35(9):937-949 [FREE Full text] [doi: 10.1093/jpepsy/jsq023] [Medline: 20392790]

Abbreviations

DHLI: digital health literacy instrument

Edited by R Kukafka; submitted 16.01.20; peer-reviewed by C Drossaert, B Nievas Soriano, A Chen; comments to author 28.02.20; revised version received 21.07.20; accepted 10.01.21; published 15.03.21.

Please cite as:
Park E, Kwon M
Testing the Digital Health Literacy Instrument for Adolescents: Cognitive Interviews
J Med Internet Res 2021;23(3):e17856
URL: https://www.jmir.org/2021/3/e17856
doi:10.2196/17856
PMID:33720031

Original Paper

Arabic Version of the Electronic Health Literacy Scale in Arabic-Speaking Individuals in Sweden: Prospective Psychometric Evaluation Study

Josefin Wångdahl[1,2], PhD; Karuna Dahlberg[3], PhD; Maria Jaensson[3], PhD; Ulrica Nilsson[2,4], PhD

[1]Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden

[2]Department of Neurobiology, Care Sciences and Society, Karolinska Institute, Stockholm, Sweden

[3]School of Health Sciences, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

[4]Department of Perioperative Medicine and Intensive Care, Karolinska University Hospital, Stockholm, Sweden

Corresponding Author:
Josefin Wångdahl, PhD
Department of Public Health and Caring Sciences
Uppsala University
Box 564
Uppsala, 751 22
Sweden
Phone: 46 702034045
Email: josefin.wangdahl@pubcare.uu.se

Abstract

Background: Health information is often communicated through the internet. It is vital for the end user to have a range of digital skills as well as understand the information to promote their health. There is a valid and reliable 8-item instrument, the Electronic Health Literacy Scale (eHEALS), that evaluates these skills. The number of Arabic-speaking people migrating to Sweden and to other parts of the world is increasing due to unstable military and political situations in their countries of origin. Poor health and limited health literacy have been described in this population in Sweden. Still, to our knowledge, an Arabic version of eHEALS has not been tested for validity or reliability. Thus, Arabic-speaking populations in Sweden cannot be included in studies measuring eHealth literacy, which does not support equal treatment in health care.

Objective: The aim of this study was to translate and adapt the original English eHEALS version into Arabic and to evaluate its psychometric properties.

Methods: The eHEALS was rigorously translated, adapted, and evaluated for content validity. We conducted prospective psychometric evaluation with natively Arabic-speaking participants living in Sweden. Construct validity, factor structure, internal consistency, and test-retest reliability were evaluated using Spearman correlation, principal component analysis, Cronbach α, and weighted quadratic Cohen κ, respectively.

Results: The study population consisted of Arabic-speaking participants (n=298; age: mean 41.8 years, SD 10.5). Construct validity was supported with weak and moderate correlations. Principal component factor analysis revealed a one-factor structure. Internal consistency was high (Cronbach α=0.92); test-retest reliability was acceptable (weighted quadratic Cohen κ=0.76). Evaluation indicated that eHealth literacy threshold values should be dichotomized (limited and sufficient) rather than trichotomized (inadequate, problematic, and sufficient).

Conclusions: The Arabic version of eHEALS, a unidimensional scale that is valid and reliable for measuring eHealth literacy among natively Arabic-speaking people in Sweden, was found to be acceptable and feasible in a general population.

(J Med Internet Res 2021;23(3):e24466) doi:10.2196/24466

KEYWORDS

eHealth; digital health literacy; eHEALS; health literacy; internet; psychometrics; evaluation, migrants, refugees, Arabic

Introduction

Arabic is 1 of 6 official languages of the United Nations and the official language of more than 20 countries. Unstable political and military situations have led to forced displacement for many people in some of these countries. In 2019, Arabic was one of the most widely spoken languages among refugees [1] (ie, someone who has been forced to flee their country because of persecution, war or violence; has a well-founded fear of persecution for reasons of race, religion, nationality, political opinion or membership in a particular social group; and that most likely cannot return home or are afraid to do so [2]) worldwide. Most refugees (approximately 6.6 million) came from Syria, which is also the most common country of origin for refugees overall. In Sweden, approximately 200,000 refugees, of which most were Syrian, speak Arabic [3]. A large number of refugees had less than good self-assessed health [4,5] and impaired psychological well-being [4-7]. Smoking, physical inactivity, and obesity (or being overweight) were also quite common [5]. At the same time, up to 73% of Arabic speaking refugees in Sweden refrain from seeking necessary health care, due to language problems, having the idea that help will not be given, or a lack of knowledge about where to go [4,5,7]. Arabic-speaking refugees from Iraq and Syria are overrepresented in the COVID-19 infections in Sweden [8], partly due to a lack of information [9].

One social determinant that may partly contribute to people's health is health literacy [10], which "is linked to literacy and entails people's knowledge, motivation and competences to access, understand, appraise and apply health information in order to make judgements and take decisions in everyday life concerning health care, disease prevention and health promotion to maintain or improve quality of life during the life course [11]." Studies show that approximately 60% of all Arabic-speaking refugees in Sweden have limited health literacy [4,5,12], a proportion that has been found among Syrian refugees in Turkey as well [13]. Among Arabic-speaking refugees in Sweden, associations have been found between limited health literacy and poor self-assessed health and between impaired psychological well-being and having refrained from seeking health care [14]. Furthermore, associations have been found between limited health literacy and poor communication, as well as with perceptions of receiving little new knowledge and help from the health examination for asylum seekers [4].

A specific form of health literacy is eHealth literacy which "is the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem [15]." The term eHealth (ie, electronic health) came into use in the year 2000, but there is no clear definition of it [16]. The internet is an important resource for health-related information and health services. To navigate and find this information requires a range of digital skills [17], which is challenging for both patients and health care staff [18]. Another potential challenge is limited language proficiency, which can result in lower understanding of health information and instructions [14,19,20]. Furthermore, there are different types of online health information of varying quality that people need to compare and evaluate. There are also rapid changes in both care routines and technology, and health information is updated frequently—yesterday's health information may not be good practice today [15] or may not even be in practice at all.

To be able to draw conclusions about limited eHealth literacy among people who have migrated from countries where another language is spoken, and consequently, may have limited competencies in the official language of the country in which they live, a reliable and valid questionnaire is needed in a language in which they are fluent. The 8-item eHealth Literacy Scale (eHEALS) measures a broad range of eHealth literacy skills [15,21] on a 5-point Likert scale (from strongly agree to strongly disagree) eHealth literacy is classified using the sum score as inadequate, 8-20; problematic, 21-26; or sufficient, 27-40 [17]. The eHEALS is available in a range of languages [15,17,22-28] but not yet in Arabic. Psychometric testing of eHEALS indicates that it is a reliable and valid instrument [15,17,27,29-31] but also that its validity requires further investigation [28] and that the newly adapted thresholds need to be confirmed [17]. Consequently, the aim of this study was to translate and adapt the eHEALS into an Arabic version and to evaluate its psychometric properties.

Methods

Study Design and Participants

This was a prospective psychometric evaluation study that included 3 phases: translation, content validity testing, and psychometric evaluation. Data collection for phases 1 and 2 took place in April 2019, and data collection for phase 3 took place from May to September 2019 [32]. The project was approved by the Regional Ethical Review Board in Stockholm, Sweden (No. 2019/5:1) and was conducted in accordance with the 1964 Declaration of Helsinki and its subsequent amendments. All participants were informed in verbal and written formats about the purpose of the study, its procedures, and that participation was voluntary and withdrawal was possible at any time. The participants were given the guarantee that their information would be kept confidential and stored securely.

Phases

Translation

The translation process was guided by the COSMIN (Consensus-Based Standards for the Selection of Health Measurement Instruments) study design checklist for Patient-reported outcome measurement instruments [33]. Permission was obtained from the creator of the eHEALS [15]. One translator, multilingual in Arabic, English, and Swedish and with Arabic as their native language, translated the original English version of eHEALS into Arabic. The translator was not a professional translator but had a high reputation in translating health surveys. Instructions to the translator were that plain language should be used and that a young person should easily be able to understand the translation; in other words, that the items should be short, easy to understand, and not contain difficult words [34]. The focus was on maintaining the meaning of the items but also to make them easy to understand and

answer for people of varying educational and health literacy levels.

The translated eHEALS (Arabic) version was compared with the Swedish version (Sw-eHEALS [17]) to ensure that the Arabic version was in line with both original English and Swedish versions. The translator and one researcher went through each item together to verify its content, the use of plain language, and similarity. Some simplifications and adjustments of the language were made.

Four individuals who were fluent in Arabic, English, and Swedish were recruited to form a committee to examine the quality of the translation [35]—2 members had previous experience in translating health survey questions to Arabic, one of whom was also an experienced educator working with individuals who have migrated to Sweden, and 2 members had experience in health communication as nurses, one of whom also worked as a research assistant in this study. The committee members worked independently and were given the Arabic translation, the original English, and the Swedish versions of eHEALS and asked to comment on spelling, grammar, whether plain language was used, and to what extent the Arabic version was consistent with the other 2 eHEALS versions. Feedback and suggestions for improvement were received by email. One of the researchers and the research assistant discussed the feedback, which resulted in some minor changes in the wording of items and response options. The English-version term *health resources* was translated as *health information* in Arabic, partly because the meaning of the word better matches and partly because it is more common in everyday Arabic. *Health resources* has also been translated as *health information* in the Swedish version of eHEALS [15,17].

The Arabic version was then tested by 6 Swedish- and Arabic-speaking laypeople (Table 1) recruited purposively and through snowball sampling [36] by one of the researchers and the research assistant. Written and verbal information about the study was provided, and a mix of gender, age, and educational levels was sought. Based on the feedback from the participants, one of the researchers and the research assistant discussed changes that might improve the Arabic version; thereafter, the research assistant made some modifications, which consisted mainly of grammar corrections and determining which word should be in bold. Finally, a professional translator was given the Arabic version with the English and Swedish versions, in order to compare the 3 versions. The translator's verdict was that the Arabic version matched the other 2 versions in terms of purpose and content.

Table 1. Demographics of the test group (n=6).

Characteristics	Value
Gender, n	
Male	3
Female	3
Age (in years), mean (range)	38 (24-52)
Country of birth, n	
Syria	2
Iraq	1
Algeria	1
Palestine	1
France	1
Number of years lived in Sweden, mean (range)	16 (1-30)
Educational levels, n	
10-12 years	1
Graduated from university	5

Content Validity

Ar-eHEALS (Multimedia Appendix 1) content validity, the degree to which the content of an instrument is an adequate reflection of the construct that it is meant to measure [34,37], was evaluated through individual interviews with the test group composed of the 6 Swedish- and Arabic-speaking laypeople who participated in phase 1. The participants were told to think aloud during as they completed the Ar-eHEALS questionnaire and indicate whether anything felt problematic. They were also asked to argue how they were reasoning when answering the each of the 8 items.

The meaning of health information was interpreted differently; some participants mentioned public health information (available on, for example, Sweden's 24-hour helpline 1177 and Family Life, a parent forum on the internet). Other participants mentioned health information in scientific articles and reports addressed to health care professionals. In some cases, health information from friends and health care was mentioned (ie, information that was not available on internet).

The importance of, and difficulty with, health information source criticism was raised in connection with 4 out of 8 items (items 1, 3, 7, 8); participants concluded that talking to and getting information directly from a doctor was best. Whether or not a

user had education in health or health care was brought up by some participants, in connection with 2 out of 8 items (items 6 and 8), as a factor regarding to what extent the user agrees with the statement (ie, assessing and using health information appropriately). Participants' reasoning about each item corresponded with the response alternative they chose and what the item aimed to measure. Five out of 8 items (items 3-7) were perceived to be closely related to the item immediately preceding the item to which they answered, which in turn made it difficult to distinguish between them. The participants mostly chose the same response alternative for the items that they thought were similar. However, all items were experienced as short and easy to understand, despite the similarities mentioned above. The response alternatives were perceived as good and no changes were made to the questionnaire after the content validity evaluation.

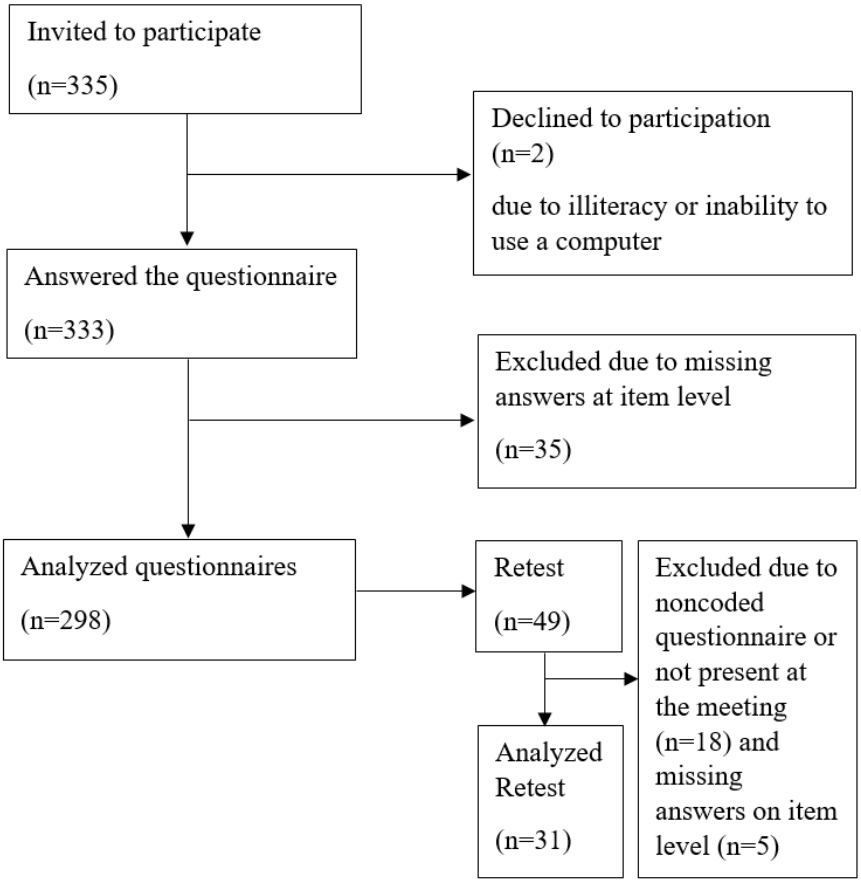
Psychometric Evaluation

Participants, Settings, and Data Collection

A sample size of 300 was considered to be appropriate [33,38]. The inclusion criteria for participation were age 18 years or older, a native speaker of Arabic, and available on the day of the data collection. The first author visited 9 different arenas: courses in civic orientation for newly arrived refugees in Sweden, fast tracks from newly arrived academics at universities, an Arabic language school, a theatre, a parent support groups, and 2 informal Arabic-language networks in order to recruit a range (gender, age, and educational levels) of participants Of the 335 people invited to take part in the study, 2 declined to participate due to illiteracy or lack of computer skills, and 35 were excluded from the study because they lacked a valid eHEALS sum score, which resulted in a study population of 298 participants (Figure 1).

To analyze test-retest reliability of Ar-eHEALS and HLS-EU-Q16, 49 participants were invited to answer the questionnaire twice with a 1-week interval. As the sample size needed for test-retest is much smaller than for testing many other forms of validity, a sample size of 25 people for the retest was considered appropriate [39]. However, in order to recruit a range and account for attrition, 49 people were asked to participate in the test-retest. To minimize dropout, participants in the test-retest groups were recruited in 2 of the groups having regular weekly meetings. To be able to compare answers from the test and the retest on individual level, the participants marked their questionnaires with a code consisting of the first 3 letters of their mothers' names and the year she was born. There was thus no need to print personal numbers and the participants remained anonymous. Of the 49 people who were asked to participate in the test-retest group, 18 people were not present at the second measurement or did not fill in the personal code to enable paring with the first questionnaire, and 5 people had at least 1 invalid Ar-eHEALS sum score; therefore, a total of 31 participants were included in the test-retest analysis (Figure 1).

Figure 1. Data collection flowchart.



Study Questionnaires and Additional Questions

Participants were given the Ar-eHEALS questionnaire, the short version of the Arabic version of the Health Literacy Survey European Questionnaire (HLS-EU-Q16) [40], questions about demographic information (age, gender, education levels, country of birth, and years lived in Sweden) and questions about health and use of the internet. The HLS-EU-Q16 consists of 16 items measuring comprehensive health literacy—perceived personal skills in finding, understanding, judging, and applying health information in order to maintain and improve health [40]. The items were answered on a 4-point Likert scale ranging from very difficult to very easy. A sum score (ranging from 0 to 16) was calculated, and self-perceived comprehensive health literacy was classified into 3 levels: 0-8, inadequate, 9-12, problematic, and 13-16, sufficient [40]. The general self-perceived health question "How do you assess your overall health status?" had "very poor, poor, fair, good, or very good" as response options [14,32,41,42]; the question "How useful is the Internet in helping you make decisions about your health?" had "not useful at all, not useful, unsure, useful, or very useful" as response options and was used to measure the usability of the internet. Importance of the internet was measured by the question "How important is it for you to be able to access health resources on the Internet?" with "not important at all, not important, unsure, important, and very important" as response options [15,17]. The question "How often do you use the Internet?" had "every day (or almost every day), several days a week, around once a week, less than day a week and never (or almost never)" as response options [17,28].

Psychometric Testing

The psychometric testing was guided by COSMIN guidelines [33,34,37,43]. Construct validity is the degree to which results from an instrument are consistent with a hypothesis [33]. In previous studies on health literacy, positive associations have been found between limited health literacy and high age [17,44-47], poor health [14,17,45,48-50], and low levels of education [17,47,51,52]. The hypotheses used in this study to evaluate the construct validity were that there is a negative correlation between Ar-eHEALS sum score and age, and there is a positive correlation between Ar-eHEALS level and education as well as between Ar-eHEALS and self-perceived general health. Other hypotheses were based on our earlier psychometric evaluation of the Swedish version of eHEALS [17], showing positive correlations between Ar-eHEALS versus interest in and level of internet use [17], the HLS-EU-Q16 sum score, and items A, K, L, and M in HLS-EU-Q16 focusing on health literacy and the Internet.

To confirm the factor structure of Ar-eHEALS principal component factoring analysis was used [53]. Cronbach α was used to assess the average correlation for the sum score of Ar-eHEALS [33].

Test-retest reliability (longitudinal reliability) was examined by calculating weighted quadratic Cohen κ coefficients [33,37].

To examine floor and ceiling effects (ie, the number of participants who choose the lowest or highest possible scores when answering), the proportion of participants who had chosen different answer alternatives was calculated. If more than 15% of a study population had chosen the lowest or highest possible score, floor or ceiling effects could be considered to be a problem [43].

The Ar-eHEALS scores were categorized according to range as inadequate (8-20), problematic (21-26), and sufficient (27-40). A dichotomization was also performed according to range: limited (8-26) and sufficient (27-40) [17].

Statistical Analysis

Data are presented as mean, standard deviation, number, percentage, or range. Spearman rank was used to analyze the correlation between the mean sum Ar-eHEALS score and the following: HLS-EU-Q16, self-perceived health, level of education, and age. A correlation coefficient magnitude between 0 and 0.1 was considered to be negligible, between 0.1 and 0.39 was considered to be weak, between 0.4 and 0.69 was considered to be moderate, between 0.7 and 0.89 was considered to be strong, and between 0.9 and 1.0 was considered to be very strong [54]. Cronbach α with a range of 0.70 to 0.95 was considered acceptable [34,55]. The weighted quadratic Cohen κ coefficient, with an accepted value of ≥0.70, was used to measure test-retest reliability [34,56]. To test thresholds, The Friedman test was used to analyze differences between Ar-eHEALS and HLS-EU-Q16 in terms of numbers of patients with inadequate, problematic, and sufficient health literacy; the Wilcoxon signed-rank test was used to analyze the threshold for limited (ie, inadequate and problematic) and sufficient health literacy. The chi-square test was used to analyze differences in gender, and the student *t* test was used to analyze differences in age. The Wilcoxon signed-rank test was used to analyze differences in educational levels, general self-perceived health, and Ar-eHEALS levels between participants with the same levels of health literacy on both the Ar-eHEALS and HLS-EU-Q16 compared with those with different levels. All data were analyzed using SPSS statistical software (version 24.0 for Windows; IBM Corp). Two-tailed *P* values <.05 were considered significant.

Results

Participants

The mean age was 41.8 years (SD 12.5), there was a higher proportion of males (180/292, 61.6%), 75% (222/296) had at least 10 years of education, and 66.4% (198/298) perceived their own general health as good or better (Table 2). On average, the participants had lived in Sweden for 9 years (range 0-38, SD 8.2). Less than half (111/289, 38.4%) had sufficient health literacy, and the mean sum score of Ar-eHEALS was 28.1 (SD 6.1). Most participants reported that they use the internet almost every day (255/297, 85.9%), that they think the internet is useful or very useful (205/298, 68.8%), and that the internet is important or very important (205/296, 69.2%). No statistically significant differences regarding gender (*P*=.19), age (*P*=.22), or education level (*P*=.93) could be found between participants who were included and those who were excluded due to missing Ar-eHEALS sum scores. Nor was any pattern of structural problems found, in terms of difficulty in responding to certain items.

Table 2. Demographics of the respondents with a valid eHEALS sum score and the test-retest group.

Characteristics	All (n=298)[a]	Test-retest group (n=31)
Gender, n (%)		
Male	112 (38.4)	9 (29.0)
Female	180 (61.6)	22 (71.0)
Age in years		
Mean (SD)	41.8 (12.5)	47.8 (10.0)
Range	21-77	30-68
Country of birth		
Syria	179 (60.1)	13 (41.9)
Iraq	65 (21.8)	11 (35.5)
Sudan	14 (4.7)	0 (0)
Other country	40 (13.4)	5 (16.1)
Number of years lived in Sweden		
Mean (SD)	9.4 (8.2)	11.2 (10.9)
Range	0-38	1-30
Highest education level, n (%)		
None	5 (1.7)	1 (3.2)
1-6 years	24 (8.1)	2 (6.5)
7-9 years	45 (15.2)	5 (16.1)
10-12 years	65 (22.0)	6 (19.4)
Graduated from university	157 (53.0)	17 (54.8)
General self-perceived health, n (%)		
Very poor	6 (2.0)	1 (3.2)
Poor	19 (6.4)	0 (0.0)
Fair	74 (24.9)	13 (41.9)
Good	127 (42.8)	12 (38.7)
Very good	71 (24.9)	5 (16.1)
HLS-EU-Q16[b], n (%)		
Inadequate	64 (22.1)	4 (13)
Problematic	114 (39.4)	10 (32)
Sufficient	111 (38.4)	17 (55)
Ar-eHEALS[c]		
Mean (SD)	28.1 (6.1)	27.8 (7.1)
Range	8-40	10-40
Frequency of internet use, n (%)		
Never	5 (1.7)	1 (3.2)
Less than 1 day 1 week	5 (1.7)	0 (0.0)
Approximately 1 day a week	7 (2.4)	0 (0.0)
Several days a week	25 (8.4)	3 (9.7)
Every day	255 (85.9)	27 (87.1)
Usability of the internet, n (%)		
Not useful at all	8 (2.7)	1 (3.2)

Characteristics	All (n=298)[a]	Test-retest group (n=31)
Not useful	21 (7.1)	3 (9.7)
Unsure	62 (20.9)	3 (9.7)
Useful	126 (42.6)	16 (51.6)
Very useful	79 (26.7)	8 (25.8)
Importance of the internet, n (%)		
Not important at all	11 (3.7)	2 (6.5)
Not important	18 (6.1)	2 (6.5)
Unsure	62 (20.9)	6 (19.4)
Important	120 (40.5)	13 (41.9)
Very important	85 (28.7)	8 (25.8)

[a]Missing responses (gender, n=6; age, n= 11, number of years lived in Sweden, n=77; highest education level, n=2; HLS-EU-Q16, n=9; frequency of internet use, n=1; importance of the internet, n=2) were not included in the denominator when calculating percentages.

[b]HLS-EU-Q16: Health Literacy Survey European Questionnaire.

[c]Ar-eHEALS: eHealth Literacy Scale.

Construct Validity

No correlation was found between the Ar-eHEALS sum score and age. A weak positive correlation was found between Ar-eHEALS and the following: education level, self-perceived health, frequency of internet use, and item A in HLS-EU-Q16. A moderate positive correlation was found between Ar-eHEALS and the following: usability of the internet; importance of the internet, HLS-EU-Q16 sum score; and items K, L, and M in HLS-EU-Q16 (Table 3).

Table 3. Spearman correlations between the Ar-eHEALS sum score and demographic characteristics, questions, and questionnaires.

Variable	Spearman ρ	*P* value
Age	–0.10	.19
Education level	0.25	<.001
Self-perceived health	0.30	<.001
Usability of the internet	0.43	<.001
Importance of the internet	0.42	<.001
Frequency of internet use	0.14	.01
HLS-EU-Q16[a] sum score	0.45	<.001
HLS-EU-Q16 item A: Find information on treatments of illnesses that concern you	0.31	<.0001
HLS-EU-Q16 item K: Judge if the information on health risks in the media is reliable	0.41	<.001
HLS-EU-Q16 item L: Decide how you can protect yourself from illness based on information in the media	0.44	<.001
HLS-EU-Q16 item M: To understand information in the media on how to get healthier	0.46	<.001

[a]HLS-EU-Q16: Health Literacy Survey European Questionnaire, 16-item.

Reliability

Internal Consistency

Factor analysis showed that the Kayser-Meyer-Olkin measure of sampling adequacy for the analysis was good (0.88, *P*<.001). Principal component analysis resulted in a one-factor solution with an initial eigenvalue of 5.0, accounting for 62.7% of the variance. The scree plot also showed a one-factor structure. All items had accepted loadings ranging from 0.63 to 0.86 (Table 4). Cronbach α=0.92 was considered acceptable as it was within the acceptable range of 0.70 to 0.95.

Table 4. Principal component analysis and weighted quadratic Cohen κ for the Arabic version of the eHealth Literacy Scale sum score or individual items.

Variable	Factor loadings	Weighted quadratic Cohen κ
eHEALS[a] sum score	N/A[b]	0.76
Item 1: I know what health resources are available on the Internet	0.63	0.78
Item 2: I know where to find helpful health resources on the Internet	0.83	0.73
Item 3: I know how to find helpful health resources on the Internet	0.86	0.83
Item 4: I know how to find helpful information[c] on the Internet	0.83	0.62
Item 5: I know how to use the health information[c] I find on the Internet to help me	0.83	0.68
Item 6: I have the skills I need to evaluate the health resources I find on the Internet	0.78	0.47
Item 7: I can tell high quality health resources from low quality health resources on the Internet	0.79	0.64
Item 8: I feel confident in using information from the Internet to make health decisions	0.77	0.47

[a]Ar-eHEALS: Arabic version of the eHealth Literacy Scale.

[b]N/A: not applicable.

[c]The term *health resources* was used instead of *health information*, which is used in the original version [15].

Test-Retest Reliability

A total of 31 participants with a mean age of 47.8 years (SD 10.0) were included in the test-retest; 29.0% (9/31) were male, 74.2% (23/31) had at least 10 years of education, and 54.8% (17/31) perceived their own general health as good or better (Table 2). The majority (17/31, 54.8%) had sufficient comprehensive health literacy, and the mean sum score of Ar-eHEALS was 27.8 (SD 7.1). Most participants reported that they use the internet almost every day (27/31, 87.1%), that they think the internet is useful or very useful (24/31, 77.4%), and that the internet is important or very important (21/31, 67.7%). Test-retest reliability of the Ar-eHEALS sum score was acceptable (Cohen κ=0.76, *P*<.001). At the items level, test-retest reliability was acceptable for 4 of 8 items (Cohen κ=0.73-0.83; Table 4).

Floor and Ceiling Effects

No floor and ceiling effects were found for the Ar-eHEALS; 3.7% (11/298) had the highest possible sum score, and 1.0% (3/298) had the lowest possible sum score.

Thresholds

When comparing numbers of participants with inadequate, problematic, and sufficient health literacy (HLS-EU-Q16) with eHealth literacy (Ar-eHEALS), there were statistically significant differences depending on which scale was used (*P*<.001). Higher proportions of inadequate (64/289, 21.5%) and problematic (114/289, 39.4%) were found for health literacy than were found for eHealth literacy; inadequate (26/289, 8.9%) and problematic (83/289, 28.7%). The opposite was found for sufficient health literacy (111/289, 38.4%) compared to sufficient eHealth literacy (180/289, 62.2%). When dichotomizing Ar-eHEALS and HLS-EU-Q16 into limited (inadequate and problematic combined) and sufficient health literacy, there was a significantly greater proportion of participants who scored the same level of health literacy on both questionnaires compared to participants who had different levels (same: 184/289, 63.7%; different: 105/289, 36.3%; *P*<.001; Table 5). There were no significant differences in age (*P*=.52), gender (*P*=.20), educational level (*P*=.77), or general self-perceived health (*P*=.11) between these 2 groups.

Table 5. Distribution of participants scoring sufficient resp. limited health literacy and e-health literacy and participants scoring different levels depending on which questionnaire.

Health literacy (HLS-EU-Q16[a])	eHealth literacy (Ar-eHEALS[b]), n (%)[c]	
	Limited	Sufficient
Limited	91 (31.5)	87 (30.1)
Sufficient	18 (6.2)	93 (32.2)

[a]HLS-EU-Q16: Health Literacy Survey European Questionnaire.

[b]Ar-eHEALS: eHealth Literacy Scale.

[c]Percentage of all participants with responses (n=289).

Discussion

Principal Findings

The results of this psychometric evaluation support the use of Ar-eHEALS to measure the self-reported eHealth literacy of Arabic-speaking people living in Sweden. The majority of the participants (179/298, 60.1%) came from Syria, which reflects the general population of the newly arrived refugees in Sweden [57]. The translation process is of great importance, especially when measuring a phenomenon such as eHealth literacy, in order to take cultural adaptation into consideration. Kalfoss [58] pointed out that some of the difficulties encountered in the translation process may be due to the forward translations where the words are translated too closely (ie, word-for-word translation, meaning that the translation focused on the words and not the meaning of the question). This was an aspect in our translation process regarding the word *health resources*. Even though a lot of effort was invested in this translation, it was obvious that the word that replaced *health resources* (ie, *health information*) was interpreted differently by the participants during the evaluation of content validity. However, *health information* is a more common term in everyday Arabic than *health resources*, and the translators indicated that *health information* is a more appropriate concept in Arabic. The creator of the original [15] was also contacted; we were informed that there have been similar problems with the concept when translating it into other languages.

Translating, adapting, and validating a questionnaire for practice or research is a time-consuming process that requires careful planning and a rigorous methodological approach to produce a reliable and valid measure of the concept of interest in the target population [58]. In the translation process, it is necessary to ensure that the forward translator has expertise in the specific topic (ie, in this case the target population) and in the construct [33]. In this study, only one forward translator was used. However, the committee that examined the quality of the translation consisted of 4 people who were multilingual in Arabic, English, and Swedish, with experience in translating health surveys, data collection in Arabic, health communication, or in education for people who have migrated. Lastly, the translation process involved laypeople and an expert for feedback on wording. This step is particularly important when an understanding of the items is vital, for example, in a questionnaire measuring literacy skills such as that measuring eHealth literacy in our study.

In the psychometric evaluation, the Ar-eHEALS and the Sw-eHEALS [17] were found to have a one-factor structure (ie, unidimensionality). The eHEALS was originally proposed to have a one-factor structure [15], which has been supported by substantial evidence which has been supported by substantial evidence irrespective of which test theory—classical or modern—was used [22,24-26,28,59-63]. The unidimensionality indicates that all the items measure a single underlying construct, in this case eHealth literacy. One can argue that confirmatory factor analysis should be more appropriate; however, principal component factoring analysis loadings are sometimes closer approximations of the true factor loadings than the loadings produced by confirmatory factor analysis. Another difference between is that confirmatory factor analysis explains a correlation matrix, whereas principal component factoring analysis identifies the major sources of variation in data. [53]. However, the sample size for this study (n=298) was appropriate, since a sample size ≥100 participants is appropriate for conducting psychometric evaluations such as internal consistency analysis [33].

According to the context-specific nature of eHealth literacy skills [15], a moderate correlation was found between eHealth literacy measured by Ar-eHEALS sum score and health literacy measured by HLS-EU-Q16 sum score and 3 of the 4 HLS-EU-Q16 items focusing on health literacy and the internet. This is in line with the findings of previous studies [17,64] showing a relationship between eHealth literacy and health literacy. There was a moderate correlation between the Ar-eHEALS and perceptions of the usability (ρ=0.43) and between the Ar-eHEALS and perceptions of the importance of using the internet (ρ=0.42), and there was a weak correlation (ρ=0.30) between Ar-eHEALS and self-perceived health, in line with findings for Sw-eHEALS [17]. A weak correlation between the Norwegian version of eHEALS and health status has been reported [64]; however, the question about self-perceived health used in this study has not yet been validated. On the other hand, it has been argued that self-perceived health is credible indicator reflecting a person's subjective general perception of health [41].

Ar-eHEALS demonstrated a high reliability (Cronbach α=0.92), which is in line with finding for other language versions of eHEALS (Cronbach $\alpha \geq 0.88$) [22,24,26,59-63] and moderate stability over time (weighted quadratic Cohen κ=0.76 coefficient for Ar-eHEALS), which was acceptable and higher than that of the Norwegian eHEALS (Cohen κ=0.61) [64] but lower than of the Swedish version of eHEALS (Cohen κ=0.86) [17].

Use of inadequate, problematic, and sufficient levels, as previously used for Sw-eHEALS [17], could not be confirmed for the Ar-eHEALS. However, when dichotomizing the threshold into sufficient and limited, the thresholds seemed to be relevant. One can argue that for a short questionnaire such as eHEALS with 8 items, 3 threshold levels is too many and may threaten the sensitivity and specificity of the questionnaire. The most important purpose must be to identify those individuals and groups who suffer from limited eHealth literacy. Nevertheless, the threshold levels for eHEALS require further evaluation in other populations and in other language versions [17].

The purpose of our study was to develop an Arabic version of eHEALS in order to include Arabic-speaking individuals living in Sweden in future eHealth literacy research. We do believe that more knowledge about associations between eHealth literacy and health outcomes, about to what extent disease prevention and health care efforts are beneficial for Arabic-speaking Swedish residents with different levels of eHealth literacy, may be important in efforts to reduce health inequalities. Since approximately 315 million people worldwide are Arabic speaking, we also think that Ar-eHEALS can be used

globally, if further validated in the specific country and context in which it is to be used.

Limitations

This study was not without limitations. One limitation is that 5 out of 6 participants in the content validity test group had university degrees. People with lower levels of education might understand and interpret the items differently. Another limitation is that it is not clear to what *health information* referred (whether it represented the Arabic or Swedish concept of health information). We suggest that in future use of the Arabic version, the concept of heath information should be specified. The sample included in this study may not be representative of all Arabic-speaking individuals in Sweden. However, the participants included were recruited from different arenas, and included different ages, genders, and levels of education. Furthermore, there were some item for which information was missing, such as for number of years lived in Sweden (77/298, 25.8% missing). One reason for this could be that those respondents were born in Sweden and therefore did not find it relevant to answer this item. Because this missing information was considered to not influence the psychometric evaluations and because all items in the Ar-eHEALS were answered, we decided to include these questionnaires in the analysis.

Conclusion

The Ar-eHEALS has rigorously and successfully been translated and culturally adapted for an Arabic-speaking population in Sweden. The psychometric testing showed that the Ar-eHEALS is valid and reliable and can be used to assess eHealth literacy among Arabic-speaking people in Sweden. Furthermore, it indicates that sum scores should be dichotomized (into sufficient and limited eHealth literacy), but further evaluation is needed.

Acknowledgments

This study is supported by *Vetenskapsrådet* (*the Swedish Research Council*; grant number 2015-02273). Special gratitude is due to the experts in the panel in phase 1 for their valuable input. We would also like to thank Anahita Amipour for helping with data collection in phase 2. Lastly, thank you to the participants of the study for their participation and help.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Arabic eHEALS version.
[DOCX File , 80 KB - jmir_v23i3e24466_app1.docx]

References

1. Global Trends: Forced Displacement in 2019. UNHCR The UN Refugee Agency. 2019. URL: https://www.unhcr.org/statistics/unhcrstats/5ee200e37/unhcr-global-trends-2019.html [accessed 2021-08-31]
2. UNHCR Resettlement Handbook and Country Chapters. UNHCR The UN Refugee Agency. 2011. URL: https://www.unhcr.org/4a2ccf4c6.html [accessed 2021-03-04]
3. Beviljade uppehållstillstånd 2009-2019 [Granted residence permits 2009-2019]. The Swedish Migration Board. 2020. URL: https://www.migrationsverket.se/download/18.2b2a286016dabb81a1854ef/1594819419165/Beviljade%20uppeh%C3%A5llstillst%C3%A5nd%202009-2019.pdf [accessed 2021-03-04]
4. Wångdahl J, Lytsy P, Mårtensson L, Westerling R. Poor health and refraining from seeking healthcare are associated with comprehensive health literacy among refugees: a Swedish cross-sectional study. Int J Public Health 2018 Apr;63(3):409-419 [FREE Full text] [doi: 10.1007/s00038-017-1074-2] [Medline: 29480325]
5. Zdravkovic S, Cuadra Björnegren C. Kartläggning av nyanländas hälsa [Mapping of the newcomers' health]. Malmö University. 2016. URL: https://muep.mau.se/handle/2043/21873 [accessed 2021-01-05]
6. Tinghög P, Malm A, Arwidson C, Sigvardsdotter E, Lundin A, Saboonchi F. Prevalence of mental ill health, traumas and postmigration stress among refugees from Syria resettled in Sweden after 2011: a population-based survey. BMJ Open 2017 Dec 29;7(12):e018899 [FREE Full text] [doi: 10.1136/bmjopen-2017-018899] [Medline: 29289940]
7. Lecerof SS, Stafström M, Emmelin M, Westerling R, Östergen P. Findings from a prospective cohort study evaluating the effects of International Health Advisors' work on recently settled migrants' health. BMC Public Health 2017 Apr 28;17(1):369 [FREE Full text] [doi: 10.1186/s12889-017-4273-0] [Medline: 28454517]
8. Veckorapport covid-19 vecka 15 publicerad 17 april 2020 [Weekly report covid-19 week 15 published 17 April]. The Public Health Agency in Sweden. 2020. URL: https://www.folkhalsomyndigheten.se/globalassets/statistik-uppfoljning/smittsamma-sjukdomar/veckorapporter-covid-19/2020/covid-19-veckorapport-vecka-15.pdf [accessed 2021-09-28]
9. Esaiasson P, Johansson B, Ghersetti M, Sohlberg J. Kriskommunikation och segregation i en pandemi. Hur boende i utsatta områden informerade sig om coronaviruset våren 2020. Arbetsrapport nr 84 [Crisis communication and segregation in a pandemic. How residents in vulnerable areas found out about the coronavirus in the spring of 2020. Work report no. 84]. Department of Journalism, Media and Communication, Gothenburg University. 2020. URL: https://www.gu.se/sites/default/files/2020-09/84.Kriskommunikation_och_segregation_i_en_pandemi_0.pdf [accessed 2021-03-04]

10. Kickbusch IS. Health literacy: addressing the health and education divide. Health Promot Int 2001 Sep;16(3):289-297 [FREE Full text] [Medline: 11509466]
11. Sørensen K, van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
12. Wångdahl J, Lytsy P, Mårtensson L, Westerling R. Health literacy among refugees in Sweden - a cross-sectional study. BMC Public Health 2014 Oct 03;14:1030 [FREE Full text] [doi: 10.1186/1471-2458-14-1030] [Medline: 25278109]
13. Assessing the health literacy and health communication needs of Syrian refugees in Turkey. World Health Organization Regional Office for Europe. URL: https://www.euro.who.int/__data/assets/pdf_file/0010/465877/Assessing-the-health-literacy-and-health-communication-needs-of-Syrian-refugees-in-Turkey-eng.pdf [accessed 2021-03-04]
14. Wångdahl J, Lytsy P, Mårtensson L, Westerling R. Health literacy and refugees' experiences of the health examination for asylum seekers - a Swedish cross-sectional study. BMC Public Health 2015 Nov 23;15:1162 [FREE Full text] [doi: 10.1186/s12889-015-2513-8] [Medline: 26596793]
15. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
16. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
17. Wångdahl J, Jaensson M, Dahlberg K, Nilsson U. The Swedish version of the Electronic Health Literacy Scale: prospective psychometric evaluation study including thresholds levels. JMIR Mhealth Uhealth 2020 Feb 24;8(2):e16316 [FREE Full text] [doi: 10.2196/16316] [Medline: 32130168]
18. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
19. Kirmayer LJ, Narasiah L, Munoz M, Rashid M, Ryder AG, Guzder J, Canadian Collaboration for Immigrant and Refugee Health (CCIRH). Common mental health problems in immigrants and refugees: general approach in primary care. CMAJ 2011 Sep 06;183(12):E959-E967 [FREE Full text] [doi: 10.1503/cmaj.090292] [Medline: 20603342]
20. Sundquist J. Migration, equality and access to health care services. J Epidemiol Community Health 2001 Oct;55(10):691-692 [FREE Full text] [doi: 10.1136/jech.55.10.691] [Medline: 11553650]
21. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? a systematic review. Knowledge Management E-Learning 2015;7(4):576-600 [FREE Full text] [doi: 10.34105/j.kmel.2015.07.038]
22. Diviani N, Dima AL, Schulz PJ. A psychometric analysis of the Italian version of the eHealth Literacy Scale using item response and classical test theory methods. J Med Internet Res 2017 Apr 11;19(4):e114 [FREE Full text] [doi: 10.2196/jmir.6749] [Medline: 28400356]
23. Tomás C, Queirós P. Analysis of the psychometric properties of the portuguese version of an eHealth literacy assessment tool. Revista de Enfermagem Referência 2014;4(2):19-28 [FREE Full text]
24. Chang A, Schulz PJ. The measurements and an elaborated understanding of Chinese eHealth literacy (C-eHEALS) in chronic patients in China. Int J Environ Res Public Health 2018 Dec 23;15(7):1553. [doi: 10.3390/ijerph15071553] [Medline: 30041420]
25. Mitsutake S, Shibata A, Ishii K, Okazaki K, Oka K. [Developing Japanese version of the eHealth Literacy Scale (eHEALS)]. Nihon Koshu Eisei Zasshi 2011 May;58(5):361-371. [Medline: 21905612]
26. Paige SR, Krieger JL, Stellefson M, Alber JM. eHealth literacy in chronic disease patients: An item response theory analysis of the eHealth literacy scale (eHEALS). Patient Educ Couns 2017 Feb;100(2):320-326. [doi: 10.1016/j.pec.2016.09.008] [Medline: 27658660]
27. Paramio PG, Almagro BJ, Hernando G, Aguaded GJI. [Validation of the eHealth Literacy Scale (eHEALS) in Spanish University Students]. Rev Esp Salud Publica 2015;89(3):329-338 [FREE Full text] [doi: 10.4321/S1135-57272015000300010] [Medline: 26388346]
28. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
29. Brørs G, Wentzel-Larsen T, Dalen H, Hansen TB, Norman CD, Wahl A, CONCARD Investigators. Psychometric properties of the Norwegian version of the Electronic Health Literacy Scale (eHEALS) among patients after percutaneous coronary intervention: cross-sectional validation study. J Med Internet Res 2020 Jul 28;22(7):e17312 [FREE Full text] [doi: 10.2196/17312] [Medline: 32720900]
30. Richtering SS, Morris R, Soh S, Barker A, Bampi F, Neubeck L, et al. Examination of an eHealth literacy scale and a health literacy scale in a population with moderate to high cardiovascular risk: Rasch analyses. PLoS One 2017;12(4):e0175372 [FREE Full text] [doi: 10.1371/journal.pone.0175372] [Medline: 28448497]
31. Koo M, Norman CD, Hsiao-Mei C. Psychometric evaluation of a Chinese version of the eHealth literacy scale (eHEALS) in school age children. Int Electron J Health Educ 2012;15(1):29-36 [FREE Full text]

32. Wangdahl JM, Dahlberg K, Jaensson M, Nilsson U. Psychometric validation of Swedish and Arabic versions of two health literacy questionnaires, eHEALS and HLS-EU-Q16, for use in a Swedish context: a study protocol. BMJ Open 2019 Sep 17;9(9):e029668 [FREE Full text] [doi: 10.1136/bmjopen-2019-029668] [Medline: 31530602]
33. Mokkink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, de Vet HCW, et al. COSMIN study design checklist for patient-reported outcome measurement instruments. COSMIN.: Amsterdam Public Health Research Institute; 2019. URL: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf [accessed 2021-09-28]
34. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
35. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993 Dec;46(12):1417-1432. [doi: 10.1016/0895-4356(93)90142-n] [Medline: 8263569]
36. Valerio MA, Rodriguez N, Winkler P, Lopez J, Dennison M, Liang Y, et al. Comparing two sampling methods to engage hard-to-reach communities in research priority setting. BMC Med Res Methodol 2016 Oct 28;16(1):146 [FREE Full text] [doi: 10.1186/s12874-016-0242-z] [Medline: 27793191]
37. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737-745. [doi: 10.1016/j.jclinepi.2010.02.006] [Medline: 20494804]
38. VanVoorhis C, Morgan B. Understanding power and rules of thumb for determining sample sizes. Tutor Quant Methods Psychol 2007;3(2):43-50 [FREE Full text] [doi: 10.20982/tqmp.03.2.p043]
39. Bujang M, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. Arch Orofac Sci 2017;12(1):1-11 [FREE Full text]
40. Sørensen K, Van den Broucke S, Pelikan JM, Fullam J, Doyle G, Slonska Z, HLS-EU Consortium. Measuring health literacy in populations: illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). BMC Public Health 2013 Oct 10;13:948 [FREE Full text] [doi: 10.1186/1471-2458-13-948] [Medline: 24112855]
41. Nielsen SS, Krasnik A. Poorer self-perceived health among migrants and ethnic minorities versus the majority population in Europe: a systematic review. Int J Public Health 2010 Oct;55(5):357-371. [doi: 10.1007/s00038-010-0145-4] [Medline: 20437193]
42. Benyamini Y. Why does self-rated health predict mortality? an update on current knowledge and a research agenda for psychologists. Psychol Health 2011 Nov;26(11):1407-1413. [doi: 10.1080/08870446.2011.621703] [Medline: 22111660]
43. de Vet HC, Terwee CB, Mokkink LB, Knol DK. Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press; 2014.
44. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
45. Palumbo R, Annarumma C, Adinolfi P, Musella M, Piscopo G. The Italian Health Literacy Project: insights from the assessment of health literacy skills in Italy. Health Policy 2016 Sep;120(9):1087-1094. [doi: 10.1016/j.healthpol.2016.08.007] [Medline: 27593949]
46. Sørensen K, Pelikan JM, Röthlin F, Ganahl K, Slonska Z, Doyle G, HLS-EU Consortium. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). Eur J Public Health 2015 Dec;25(6):1053-1058 [FREE Full text] [doi: 10.1093/eurpub/ckv043] [Medline: 25843827]
47. Paasche-Orlow MK, Parker RM, Gazmararian JA, Nielsen-Bohlman LT, Rudd RR. The prevalence of limited health literacy. J Gen Intern Med 2005 Feb;20(2):175-184 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.40245.x] [Medline: 15836552]
48. Hälleberg Nyman M, Nilsson U, Dahlberg K, Jaensson M. Association between functional health literacy and postoperative recovery, health care contacts and health related quality of life among patients undergoing day surgery. JAMA Surg 2018 Aug 01;153(8):738-745 [FREE Full text] [doi: 10.1001/jamasurg.2018.0672] [Medline: 29710226]
49. Easton P, Entwistle VA, Williams B. Health in the 'hidden population' of people with low literacy. a systematic review of the literature. BMC Public Health 2010 Aug 05;10:459 [FREE Full text] [doi: 10.1186/1471-2458-10-459] [Medline: 20687946]
50. Levin-Zamir D, Baron-Epel OB, Cohen V, Elhayany A. The association of health literacy with health behavior, socioeconomic indicators, and self-assessed health from a national adult survey in Israel. J Health Commun 2016;21(sup2):61-68. [doi: 10.1080/10810730.2016.1207115] [Medline: 27669363]
51. Almaleh R, Helmy Y, Farhat E, Hasan H, Abdelhafez A. Assessment of health literacy among outpatient clinics attendees at Ain Shams University Hospitals, Egypt: a cross-sectional study. Public Health 2017 Oct;151:137-145. [doi: 10.1016/j.puhe.2017.06.024] [Medline: 28800559]
52. Ng E, Omariba DWR. Immigration, generational status and health literacy in Canada. Health Educ J 2013 Nov 27;73(6):668-682 [FREE Full text] [doi: 10.1177/0017896913511809]

53. de Winter JCF, Dodou D. Common factor analysis versus principal component analysis: a comparison of loadings by means of simulations. Commun Stat Simul Comput 2015 Feb 25;45(1):299-321. [doi: 10.1080/03610918.2013.862274]
54. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. Anesth Analg 2018 May;126(5):1763-1768. [doi: 10.1213/ANE.0000000000002864] [Medline: 29481436]
55. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/bf02310555]
56. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
57. Parkvall M. Arabiska Sveriges näst största modersmål [Arabic Sweden's second largest mother tongue]. Svenska Dagbladet. 2018. URL: https://www.svd.se/arabiska-sveriges-nast-storsta-modersmal [accessed 2021-01-05]
58. Kalfoss M. Translation and adaption of questionnaires: a nursing challenge. SAGE Open Nurs 2019 Jan 23;5:2377960818816810 [FREE Full text] [doi: 10.1177/2377960818816810] [Medline: 33415214]
59. Chung S, Nahm E. Testing reliability and validity of the eHealth Literacy Scale (eHEALS) for older adults recruited online. Comput Inform Nurs 2015 Apr;33(4):150-156. [doi: 10.1097/CIN.0000000000000146] [Medline: 25783223]
60. Chung S, Park BK, Nahm E. The Korean eHealth Literacy Scale (K-eHEALS): reliability and validity testing in younger adults recruited online. J Med Internet Res 2018 Apr 20;20(4):e138 [FREE Full text] [doi: 10.2196/jmir.8759] [Medline: 29678800]
61. Lin C, Broström A, Griffiths MD, Pakpour AH. Psychometric evaluation of the Persian eHealth Literacy Scale (eHEALS) among elder Iranians with heart failure. Eval Health Prof 2019 Feb 11:163278719827997. [doi: 10.1177/0163278719827997] [Medline: 30744419]
62. Juvalta S, Kerry MJ, Jaks R, Baumann I, Dratva J. Electronic health literacy in Swiss-German parents: cross-sectional study of eHealth literacy scale unidimensionality. J Med Internet Res 2020 Mar 13;22(3):e14492 [FREE Full text] [doi: 10.2196/14492] [Medline: 32167476]
63. Ma Z, Wu M. The psychometric properties of the Chinese eHealth literacy scale (C-eHEALS) in a Chinese rural population: cross-sectional validation study. J Med Internet Res 2019 Oct 22;21(10):e15720 [FREE Full text] [doi: 10.2196/15720] [Medline: 31642811]
64. Neter E, Brainin E, Baron-Epel O. The dimensionality of health literacy and eHealth literacy. Eur Health Psychologist 2015;17(6):275-280 [FREE Full text]

Abbreviations

Ar-eHEALS: Arabic version of eHEALS
eHEALS: eHealth Literacy Scale
HLS-EU-Q16: Health Literacy Survey European Questionnaire–16
Sw-eHEALS: Swedish version of eHEALS

Edited by G Eysenbach; submitted 21.09.20; peer-reviewed by P Gaspar, S Ghaddar; comments to author 24.11.20; revised version received 15.01.21; accepted 29.01.21; published 22.03.21.

Please cite as:
Wångdahl J, Dahlberg K, Jaensson M, Nilsson U
Arabic Version of the Electronic Health Literacy Scale in Arabic-Speaking Individuals in Sweden: Prospective Psychometric Evaluation Study
J Med Internet Res 2021;23(3):e24466
URL: https://www.jmir.org/2021/3/e24466
doi:10.2196/24466
PMID:33749614

Original Paper

Contribution of Free-Text Comments to the Burden of Documentation: Assessment and Analysis of Vital Sign Comments in Flowsheets

Zhijun Yin[1,2], MSc, PhD; Yongtai Liu[2], MSc; Allison B McCoy[1], PhD; Bradley A Malin[1,2,3], PhD; Patricia R Sengstack[4], DNP, RN

[1]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

[2]Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, United States

[3]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, United States

[4]School of Nursing, Vanderbilt University, Nashville, TN, United States

Corresponding Author:
Zhijun Yin, MSc, PhD
Department of Biomedical Informatics
Vanderbilt University Medical Center
2525 West End Ave., Suite 1475
Nashville, TN, 37203
United States
Phone: 1 615 936 3690
Email: zhijun.yin@vanderbilt.edu

Abstract

Background: Documentation burden is a common problem with modern electronic health record (EHR) systems. To reduce this burden, various recording methods (eg, voice recorders or motion sensors) have been proposed. However, these solutions are in an early prototype phase and are unlikely to transition into practice in the near future. A more pragmatic alternative is to directly modify the implementation of the existing functionalities of an EHR system.

Objective: This study aims to assess the nature of free-text comments entered into EHR flowsheets that supplement quantitative vital sign values and examine opportunities to simplify functionality and reduce documentation burden.

Methods: We evaluated 209,055 vital sign comments in flowsheets that were generated in the Epic EHR system at the Vanderbilt University Medical Center in 2018. We applied topic modeling, as well as the natural language processing Clinical Language Annotation, Modeling, and Processing software system, to extract generally discussed topics and detailed medical terms (expressed as probability distribution) to investigate the stories communicated in these comments.

Results: Our analysis showed that 63.33% (6053/9557) of the users who entered vital signs made at least one free-text comment in vital sign flowsheet entries. The user roles that were most likely to compose comments were registered nurse, technician, and licensed nurse. The most frequently identified topics were the notification of a result to health care providers (0.347), the context of a measurement (0.307), and an inability to obtain a vital sign (0.224). There were 4187 unique medical terms that were extracted from 46,029 (0.220) comments, including many symptom-related terms such as “pain,” “upset,” “dizziness,” “coughing,” “anxiety,” “distress,” and “fever” and drug-related terms such as “tylenol,” “anesthesia,” “cannula,” “oxygen,” “motrin,” “rituxan,” and “labetalol.”

Conclusions: Considering that flowsheet comments are generally not displayed or automatically pulled into any clinical notes, our findings suggest that the flowsheet comment functionality can be simplified (eg, via structured response fields instead of a text input dialog) to reduce health care provider effort. Moreover, rich and clinically important medical terms such as medications and symptoms should be explicitly recorded in clinical notes for better visibility.

(J Med Internet Res 2021;23(3):e22806) doi:10.2196/22806

KEYWORDS

electronic health system; documentation burden; flowsheets; content analysis; vital sign comments; free text

Introduction

Background and Motivations

Electronic health record (EHR) systems have been widely adopted in clinical settings over the past decade [1]. These systems have provided many benefits that include, but are not limited to, improving quality of care [2], reducing prescription errors [3], and facilitating biomedical research [4]. Despite such benefits, documentation burden has been recognized as a negative artifact of adopting EHR systems. For instance, it was shown that primary care clinicians spent more than 50% of their time in front of an EHR system, thus reducing their time in interactions with patients [5]. In another study, it was reported that ophthalmologists spent approximately 3.7 hours per day using EHRs [6]. It has also been shown that nurses, one of the largest EHR system users, enter approximately 640 flowsheet data entries during a 12-hour shift, nearly one data point every minute in acute care [7]. In addition, in a web-based survey conducted in the Nursing Quality and Care Forum, 78% of the participants confirmed that documentation in EHRs is time-consuming and difficult to complete, and 68% suggested that such documentation contributed little value to patient care [8].

Documentation burden originates from various factors, such as the complex functionalities of EHR systems (which itself is partially due to increasingly sophisticated health care routines), increase in the amount of data being collected, and the challenge of prioritizing the information scattered in different locations in an EHR system [9]. The US Department of Health and Human Services has released strategies to reduce the burden of using health information technology (and EHRs in particular) [10], noting that the causes of documentation burden are many and complex and must be addressed on several levels by EHR vendors, regulatory agencies, insurers, and health care organizations themselves. In particular, one of the proposed strategies is to simplify documentation requirements for evaluation and management by streamlining Medicare Physician Fee Schedule final rules.

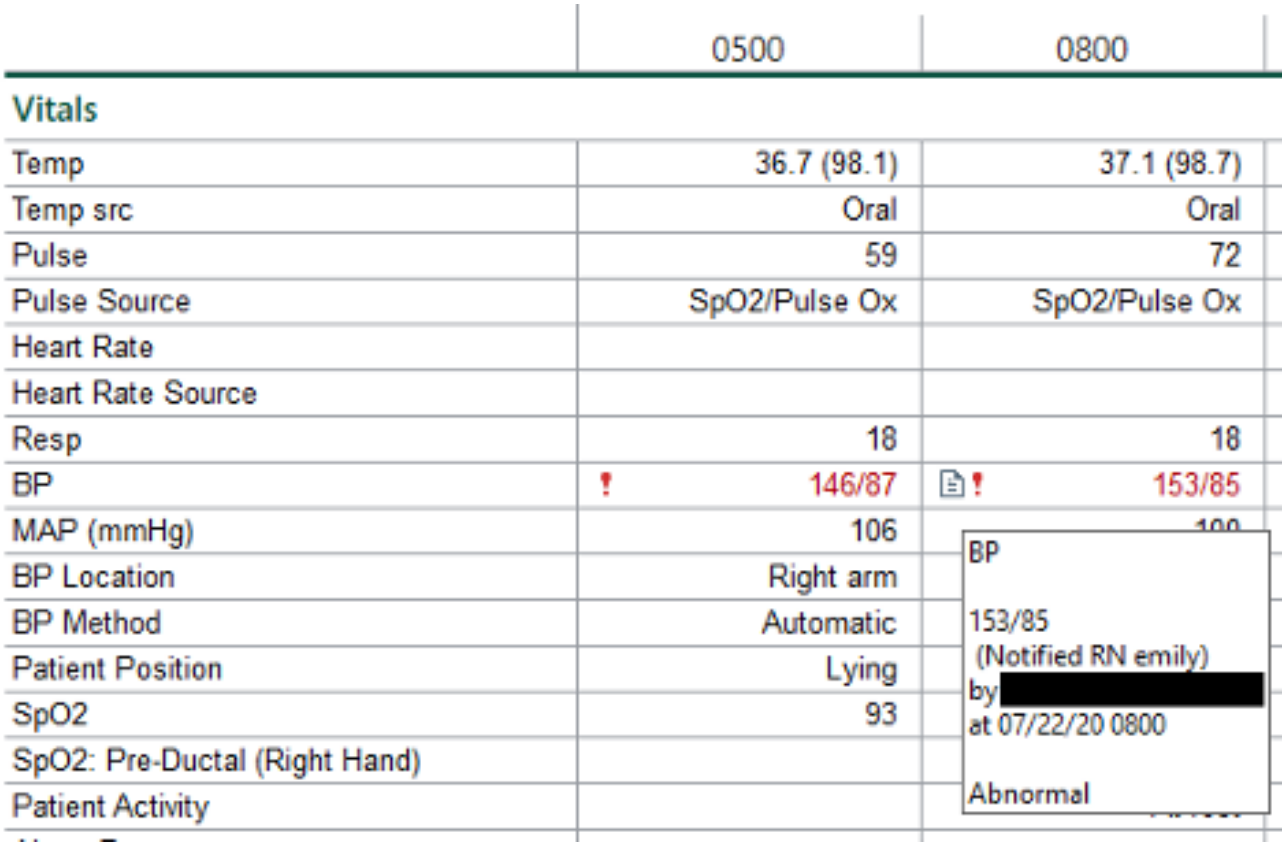
In addition to policy changes, it has been suggested that alternative recording strategies could reduce documentation burden. In one study, a smartwatch app with voice recognition was designed to help nurses record discussions during patient care, which could subsequently be uploaded to the EHR system [11,12]. A more recent study suggested that clinical documentation based on a collaborative wiki could provide opportunities to reduce documentation burden [13]. Other proposals include using artificial intelligence apps for auto-generation (eg, for treatment planning or summarization for radiation oncology [14]) or motion sensors and cameras to automatically populate EHR data (eg, in emergency care [15]). However, these approaches are limited in that most are in a prototype phase and are unlikely to be ready for implementation in the near future. Although it has been shown that using medical scribes (individuals who specialize in transcribing information during encounters into EHRs in real time) can reduce the documentation burden for physicians [16,17], scribes require a significant amount of training and clear coordination with physicians. In addition, the presence of a scribe might cause uncomfortable conversations during physician-patient encounters.

An alternative solution, with the potential for an immediate effect, is to customize the functionalities of an existing EHR system. For instance, it was shown that turning off certain interruptive notifications could help reduce EHR alert fatigue [18,19]. However, before doing this, it is necessary to investigate the functionality that is going to be customized to minimize negative consequences. In addition, organizations are beginning to examine free-text comments in EHRs, particularly those found in nursing flowsheets where the intent is to provide a place for succinct and standard responses.

Free-Text Comments in Flowsheets

Flowsheets are standardized tools in EHR systems that are helpful in documenting longitudinal patient information (eg, assessments, observations, and routine care) in a grid-type format [20]. In each flowsheet entry, a health care provider can enter values into a cell from provided lists or types in numerical values such as blood pressure (BP) or temperature. Additional free-text comments can be entered into a flowsheet cell, but this is not mandatory. By default, the comments (if any) are hidden behind an icon within the flowsheet entry. Health care providers can review a comment by clicking or hovering over the icon to open the comment display dialog. Figure 1 depicts a screenshot of a vital sign flowsheet with comments entered for BP. Although the flowsheet comments are optional, some health care providers find them useful and make an extra effort to provide them [21]. However, comments may introduce a documentation burden stemming from limitations in the existing EHR functionality. Given that flowsheet comments are made accessible in a nonobvious manner, we believe that their content can be leveraged to design more effective strategies for efficiently recording them.

Figure 1. Example of a vital sign comment that is entered in the Epic system. © 2020 Epic Systems Corporation.



Research Objectives

In this study, we seek to investigate the nature of flowsheet comments and their contribution to the documentation burden. In particular, we focused on the vital signs in flowsheets and extracted all their related comments written in 2018 in Epic, the EHR system that is in use at the Vanderbilt University Medical Center (VUMC). Specifically, we investigated the following research questions (RQs):

- RQ1: How often are the free-text comments in vital sign flowsheet entries made and by whom?
- RQ2: What are the general topics communicated in these comments?
- RQ3: Are there any specific medical terms mentioned in these comments?

Investigating the first question provides insight into how often health care providers use the flowsheet comment functionality. Answering the latter 2 questions provides insight into potential improvements to this functionality in an EHR system. Without an understanding of how often this feature is used and the purpose it serves, it is difficult to identify potential improvements to reduce the need to add documentation beyond the expected and standard response in a flowsheet cell. If medical terms or concerns are being added to essentially hidden flowsheet comments, organizations can identify system usability enhancements to better capture patient issues that need to be addressed.

Methods

Data Preparation

In this study, we collected all vital signs and their comments (if present) that were recorded in flowsheets between January 1, 2018, and December 31, 2018, at VUMC. We focused on 5 specific vital signs that are commonly collected for routine clinical use in both the inpatient and outpatient settings: body temperature (*Temp*), *BP*, oxygen saturation (SpO_2), pulse rate (*Pulse*), and respiration rate (*Resp*). For each vital sign flowsheet entry, we collected the user ID, user role, documented time, and the free-text comment entered. Our study did not involve any patients and was designated as exempt from human subject research under a VUMC Internal Review Board protocol. Multimedia Appendix 1 presents the number and percentage of other types of vital signs.

Commenting Statistics and Temporal Trend

To investigate RQ1, for each vital sign, we captured the total number of unique users who entered at least one value in the flowsheet *(total users*), the number of unique users who made at least one comment *(users commenting*), the total number of flowsheet entries *(total entries*), the number of entries with comments *(entries with comments*), and the median number of words in the comments (*comment length*). We also showed the temporal trend by illustrating the number of comments of each vital sign that were generated weekly in 2018. In addition, we counted the number of comments per user role and ranked them based on their comment volume in descending order. We report the top-ranked roles that together generate at least 90% of all the comments.

Topic Modeling

To gain insights into what was communicated in these comments, we had to rely on an efficient method to summarize such a large volume of free texts. Topic modeling is a computational method for discovering the latent *topics* that occur in a collection of documents. To apply this technique, we first manually cleaned the comments by replacing commonly misspelled words with canonical representations. For example, we replaced *rnnotifed* with *rn notified.* After data cleaning, we applied latent Dirichlet allocation (LDA), specifically its implementation in the Gensim Python package (version 3.8.0), to identify topics. LDA is a common topic modeling technique in natural language processing to infer 2 distributions from a large number of documents. The first distribution describes the probability that a topic is sampled to form a document. The second distribution describes the probability that a term is sampled from a topic. We used the first distribution to determine the popular topics mentioned in vital sign comments and the second distribution to explain what a specific topic is talking about.

As LDA is an unsupervised learning method, we applied the coherence score (specifically, C_v with a default sliding window of 110) to optimize the number of topics. The coherence score measures the extent to which the most relevant terms (with the highest probabilities) in a topic coexist with each other in either an external data set or the documents that are applied to train topic modeling. The higher the coherence score, the more interpretable the topics. In this study, we treated each vital sign as a single document and trained LDA models for 2 to 30 topics (with a step size of 1) using all the vital sign comments. Each candidate model was trained 10 times based on a different random seed. Although the best practice is to select the model that achieves the largest coherence score [22], there may be multiple LDA models that achieve coherence scores that are not significantly different from each other. As such, we empirically chose a model from these candidates that has (1) a large average coherence score, which leads to high interpretability; (2) a small SD, which tends to generate a stable model; and (3) a small number of topics, which reduces the chance of overlaps between topics.

Medical Terms Extraction

LDA is often effective at characterizing what is generally discussed in documents because it estimates the probabilities from term frequency (where a higher frequency indicates a larger probability). However, this technique is limited in that it is not oriented to represent detailed information, which is particularly a concern when relevant terms are rare. On the basis of this fact, we further applied Clinical Language Annotation, Modeling, and Processing (CLAMP, version 1.6.0), a toolkit that incorporates named-entity recognition algorithms, to identify medical terms with respect to 3 categories, that is, problems, treatments, and laboratory tests, as defined in CLAMP [23]. Examples of such terms are presented in the Results section. We use these medical terms to supplement the topics to gain a better understanding of the content of comments.

Results

Summary Statistics

During the 1-year study period (2018), there were a total of 209,055 free-text comments entered into flowsheets to further explain the data values entered for vital signs. Table 1 shows the basic statistics of the collected data. It can be seen that 63.33% (6053/9557) of the users who entered any vital signs made at least one comment in the vital sign flowsheet entries. Although BP received the second smallest number of flowsheet entries, it had the largest proportion of users who made comments and the largest number (proportion) of entries with comments. Similarly, Temp had the smallest number of total flowsheet entries but the second largest proportion of entries with comments. In contrast, Pulse and Resp had the smallest number (proportion) of users who made comments and the smallest proportion of entries with comments. Among these 5 vital signs, SpO_2 had the largest number of vital sign entries. In total, 0.69% (209,055/29,995,045) of the vital sign entries received additional comments.

Table 1. Data summary statistics[a].

Type	Users commenting, n (%)	Total users, n	Entries with comment, n (%)	Total entries, n
BP[b]	4733 (52.25)	9058	107,413 (2.04)	5,268,477
Pulse	3066 (33.76)	9081	16,814 (0.21)	7,898,699
Resp[c]	2346 (28.85)	8132	10,883 (0.18)	5,994,777
SpO_2[d]	3614 (44.52)	8118	38,819 (0.55)	7,058,507
Temp[e]	3631 (44.08)	8238	35,124 (0.93)	3,774,585
Total	6053 (63.33)	9557	209,055 (0.69)	29,995,045

[a]Note that the users can document multiple types of vital signs. As a result, the total number of unique users is the union, as opposed to the sum, of the set users associated with a vital sign.

[b]BP: blood pressure.

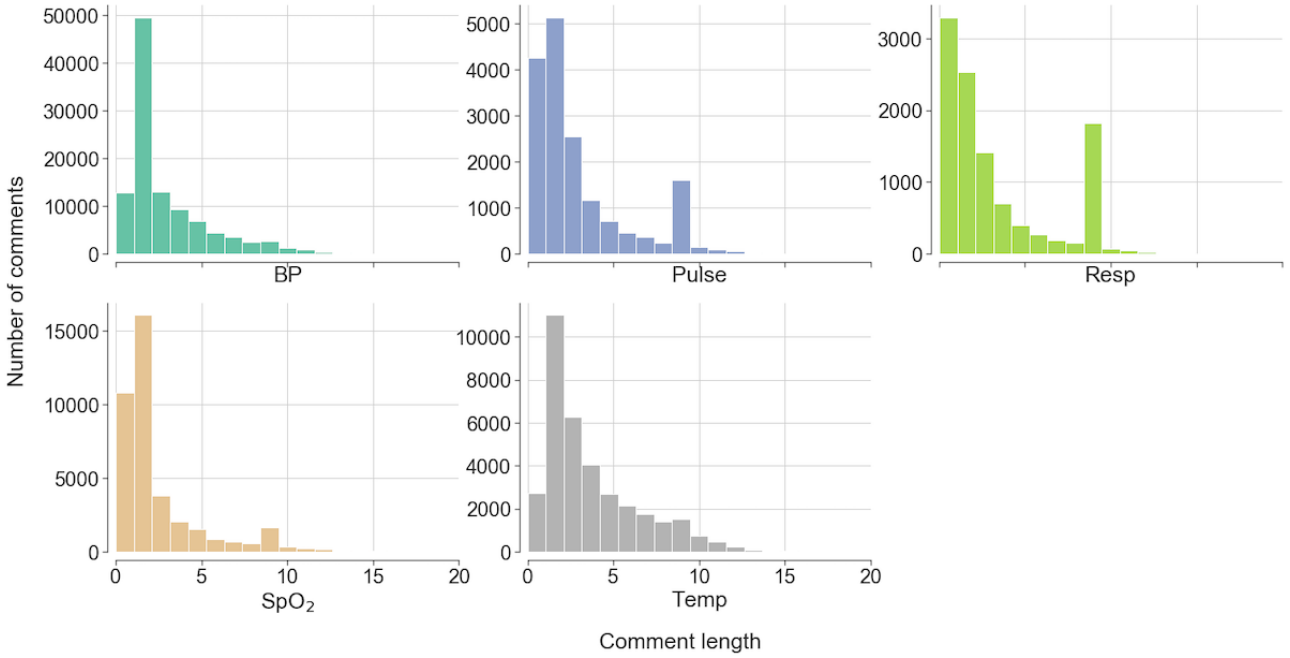
[c]Resp: respiration rate.

[d]SpO_2: oxygen saturation.

[e]Temp: body temperature.

Figure 2 shows the histogram of the comment length (the number of words) for each vital sign. From the figure, it can be seen that for almost all the comments, the number of words was less than 15. Although most comments were short (with a median number of words of 2), the number of all the words used in these comments was still 697,340, owing to the large data volume.

Figure 2. Histogram of comment length for each type of vital sign. BP: blood pressure; Pulse: pulse rate; Resp: respiration rate; SpO_2: oxygen saturation; Temp: body temperature.



Comments Stratified by User Role

Figure 3 depicts the user roles that generated at least 90% of the comments for each vital sign. It can be seen that the user roles that were most likely to compose comments were registered nurse, technician, and licensed nurse. Although medical assistant and nursing student were among the top-ranked user roles, they generated a substantially smaller number of comments. It can also be seen that the user role registered nurse generated the largest number of comments for Temp, SpO_2, Resp, and Pulse, whereas the user role technician generated the largest number of BP comments. The user role licensed nurse generated the second largest number of SpO_2 comments.

Figure 3. User roles with the largest number of comments per vital sign. Only the user roles that together generated at least 90% of the comments for each vital sign are shown. Registered Nurse and Technician are two user roles that generated the largest number of vital sign comments. BP: blood pressure; Pulse: pulse rate; Resp: respiration rate; SpO_2: oxygen saturation; Temp: body temperature.

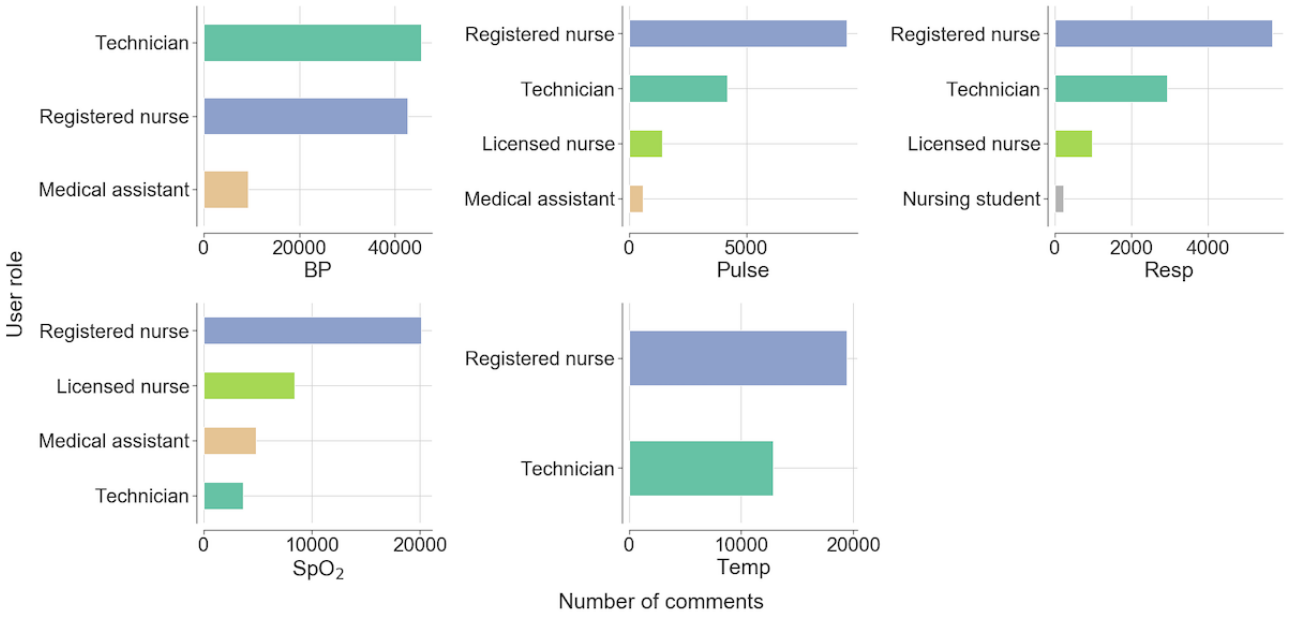
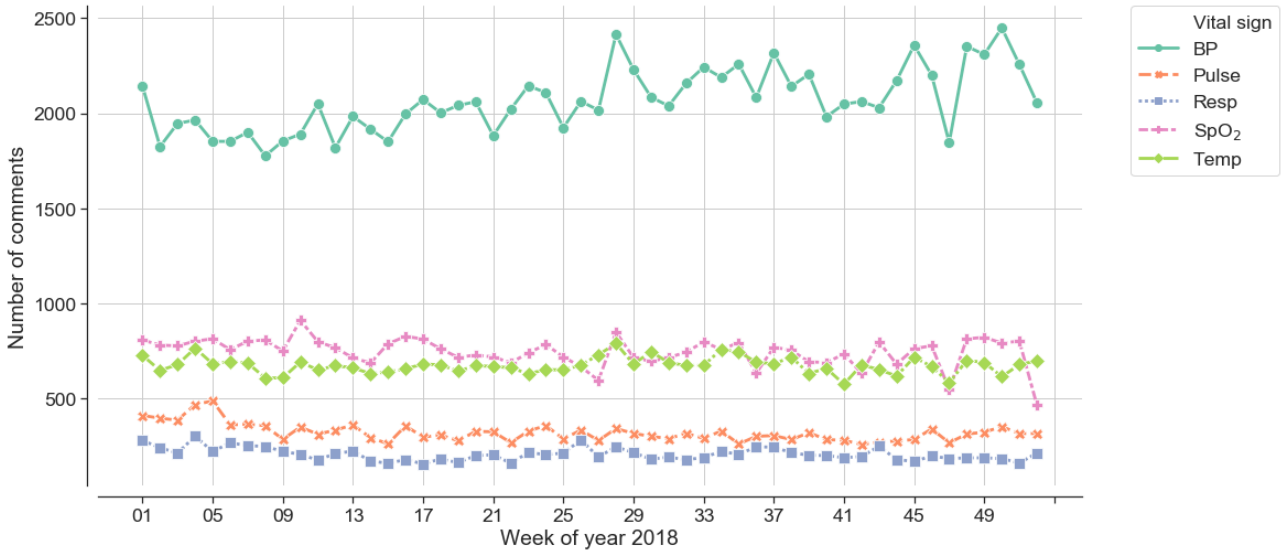


Figure 4 shows the number of comments for each vital sign in each week in 2018. Although the number of BP comments had a slightly increasing trend, the other 4 vital signs had a relatively constant number of comments for each of the 52 weeks. This suggested that the commenting phenomenon was quite stable in this clinical setting.

Figure 4. The temporal patterns of the number of comments for each type of vital sign. BP: blood pressure; Pulse: pulse rate; Resp: respiration rate; SpO_2: oxygen saturation; Temp: body temperature.



Topic Analysis

On the basis of our criteria, we empirically generated 13 topics from the flowsheet comment (see Multimedia Appendix 1 for details on how 13 topics were selected). Table 2 shows the topics, their most relevant terms, and the probability distribution. The relevant terms were selected based on their probability rank (in descending order) within a topic. For example, “notify,” “uto” (unable to obtain), “fussy,” and “move” were the most relevant terms in topic T8, indicating that a related measurement might not be obtained. Owing to the overlap between topics (eg, though a manual review), we further categorized these topics into 5 groups by examining the topic words and the associated comments (eg, examining the comments with the largest distribution of a particular topic). The percentage of each group was calculated by summing the percentage of each topic within the group. From the table, it can be seen that most comments communicated the notification of a result to health care providers (0.347), the context of a measurement (0.307), and an inability to obtain a vital sign (0.224). The other 2 topics corresponded to the measurement method (0.071) and simultaneous filling (0.051).

Table 2. Topics generated from vital sign comments[a].

Label for group and topic	Most relevant words	Topic distribution (probability)	Group distribution (probability)
Notification			0.347
T1	notify, rn[b], nurse, lie, standing, manually, informed, <NAME>, trach, <NAME>	0.106	
T3	nurse, notify, call, high, pressure, heat, primary, hfov[c], report, warmer	0.090	
T9	notify, rn, nurse, elevated, <NAME>, abnormal, supine, <NAME>, <NAME>, <NAME>	0.076	
T6	notif, team, md[d], bedside, notifed, order, cct[e], page, monitor, np[f]	0.075	
Context			0.307
T13	room, air	0.092	
T4	arm, pt[g], nc[h], blanket, place, warm, apply, hugger, bair, baby	0.080	
T12	post, bath, temp[i], stand, sit, min[j], provider, eoi[k], inform, care	0.075	
T7	patient, give, sleeping, tylenol, pain, med, state, liter, due, floor	0.060	
Unable to obtain			0.224
T8	patient, uto, fussy, move, crying, kicking, agitate, screaming, unit, moving	0.084	
T2	patient, refuse, vital, sleep, asleep, time, mom, awake, defer, request	0.072	
T11	patient, move, cry, upset, good, attempt, uto, obtain, kick, uta	0.068	
Measure method			0.071
T10	manual, cuff, bp[l], unable, nurse, check, arm, temp, read, recheck	0.071	
Simultaneous filing			0.051
T5	datum, user, filing, simultaneous, previous, doppler, predose, cchd[m], unnotifed, present	0.051	

[a]Topic T13 only has 2 words with positive probabilities, whereas the probability of all the other words was zero and are thus not displayed. The names in T1 and T9 are replaced with <NAME> for anonymity.

[b]rn: registered nurse.

[c]hfov: high-frequency oscillatory ventilation.

[d]md: doctor of medicine.

[e]cct: critical care team.

[f]np: nurse practitioner.

[g]pt: patient.

[h]nc: nasal cannula.

[i]temp: body temperature.

[j]min: minute.

[k]eoi: evidence of insurability.

[l]bp: blood pressure.

[m]cchd: critical congenital heart disease.

To better understand each topic, we showed the comment samples, their dominant topics (the topic with the largest probability), and topic percentages in Table 3. It should be noted that the names of health care providers mentioned in some notification-related samples were replaced with <NAME> for anonymity. Although the meanings of the samples were straightforward and clearly linked to the associated topic groups, there are still several observations that we want to highlight here. First, despite the short length, the content of comment samples contained rich information, some of which was beyond

the vital signs themselves. For example, some samples in the context and notification topic groups included information regarding medications (eg, nitro paste and BP meds). This confirmed the necessity of conducting a further medical term analysis to obtain more insights into this type of information. Second, the unable to obtain topic group mainly documented why a measurement was not obtained. Finally, after a close examination of the comments with T10 (simultaneous filing topic group) as the dominant topic, we found that this topic might refer to a conflicting input of a vital sign between an automatic interface and a health care provider.

Table 3. Examples of comments for each topic[a].

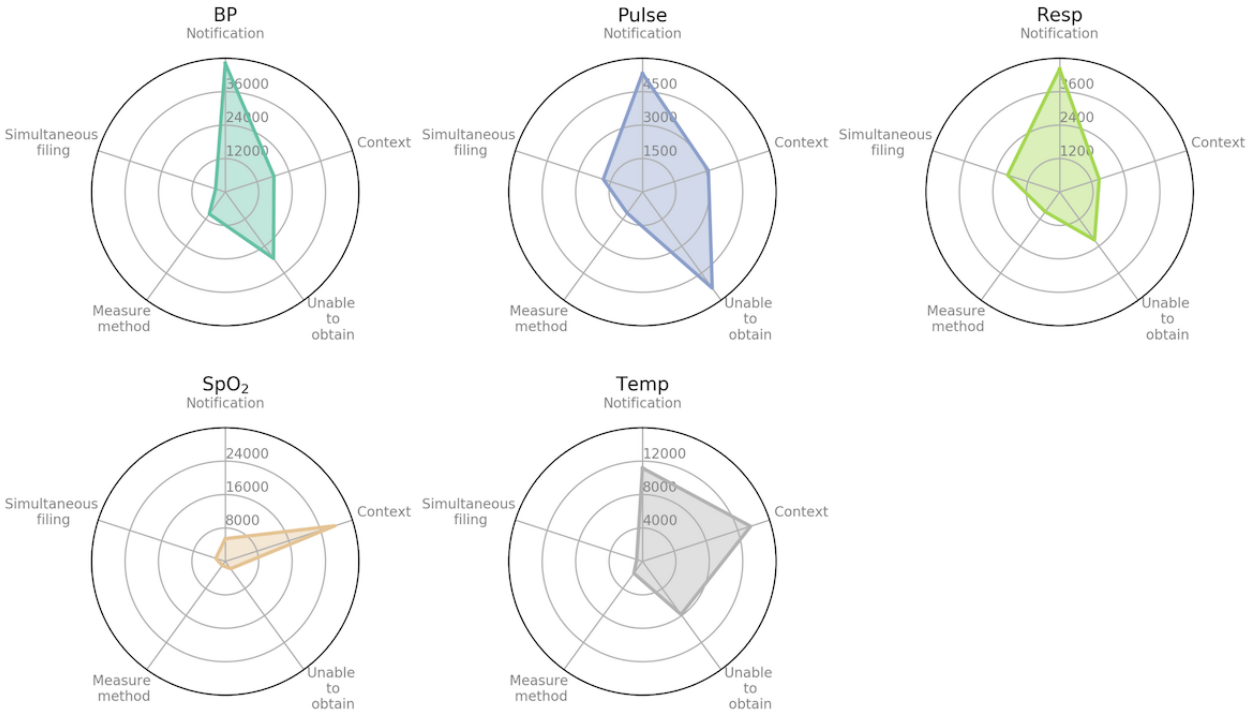
Topic group	Topic	Topic distribution (probability)
Notification		
"Notified <NAME>, RN and <NAME>, MSN"	T1	0.182
"Nurse notified that pressure is high"	T3	0.139
"<NAME>, RN and Professor <NAME> notified"	T9	0.147
"Paged neuro stroke team about BP, team ordered nitro paste"	T6	0.204
Context		
"Room air, baseline oxygen sat 87% on room air preop"	T13	0.135
"Eating cold food and has heating pad under back/arm area"	T4	0.183
"Transfusion ended. No s/sx of blood tXXXransfusion reaction"	T12	0.170
"Patient has not taken her BP[b] meds today"	T7	0.135
Unable to obtain		
"uto, patient fussy, moving"	T8	0.143
"Mom refused vitals, requests patient not be disturbed until wakes"	T2	0.184
"Patient upset, no BP obtained, multiple attempts"	T11	0.151
Measure method		
"Manual BP (arm measured 31 cm, used adult size cuff)"	T10	0.181
Simultaneous filing		
"Paced simultaneous filing. User may not have seen previous data"	T11	0.161

[a]The dominant topic (eg, the topic with the largest probability) in each topic group is shown in the Topic column, and the corresponding probability is shown in the Topic Distribution column.

[b]BP: blood pressure.

Figure 5 shows how each topic group is represented in each vital sign. Specifically, we used the dominant topic to determine which topic group a comment is assigned to and then counted the number of comments in each topic group for each vital sign. For example, all the 4 comment samples under *Notification* in Table 3 were assigned to this topic group because their dominant topics (T1, T3, T6, and T9) belonged to *Notification* (Table 2). It can be seen that some topic groups are specific to certain vital signs. For example, *notification* was the dominant topic group in BP and Resp, whereas *context* was the dominant topic group in SpO_2. In addition, Pulse had *notification* and *unable to obtain* as dominant topic groups, whereas temp had *context* and *notification* as dominant topic groups.

Figure 5. Distribution of topic groups in each vital sign. Each comment is assigned to the topic group that its dominant topic belongs to. BP: blood pressure; Pulse: pulse rate; Resp: respiration rate; SpO_2: oxygen saturation; Temp: body temperature.



Medical Term Extraction

We applied CLAMP to generate 4187 unique medical terms or phrases that were found in 22.02% (46,029/209,055) of all vital sign comments. Of these comments, 7.66% (16,023/209,055) contained treatment-related keywords, 8.66% (18,095/209,055) contained test-related keywords, and 7.16% (14,978/209,055) contained problem-related keywords. Multimedia Appendices 2-4 depict the word clouds of the medical terms in each category. The terms *bp* and *fussy* covered 30.47% (5814/19,080) and 24.68% (3941/15,966) of all the test- and problem-related medical terms, respectively. Given their dominance, they are not reported in Multimedia Appendices 2 and 3. From the figure, we can see that test-related medical terms were mainly related to vitals, which intuitively makes sense because this study focused on vital sign comments. However, there were other tests, such as magnetic resonance imaging, which were mentioned in these comments. The problem-related terms included many symptom-related terms such as “pain,” “upset,” “dizziness,” “coughing,” “anxiety,” “distress,” and “fever.” Treatment-related terms included “tylenol,” “anesthesia,” “cannula,” “oxygen,” “motrin,” “rituxan,” “bair hugger blanket,” and “labetalol.”

Discussion

Principal Findings

This study has several notable findings. First, we found that 63.33% (6053/9557) of the EHR users who recorded vital signs in flowsheets also entered at least one comment, and most of these users were either (registered) nurses or technicians. Although only 0.69% (209,055/29,995,045) of the 5 types of vital sign entries received comments, there were approximately 210,000 comments. Furthermore, our topic analysis showed that these comments mainly corresponded to how the vital signs were measured, any issues encountered while taking the vital signs, or notifications to health care providers. Further inspection of the medical terms indicated that there were still many test-, problem-, and treatment-related data that were recorded in these comments. These nurses and technicians clearly felt the need to capture more information than the simple numeric value that the flowsheet required, as documented in a study that examined the flowsheet comments for 201 patients who experienced cardiac arrest [21]. However, despite its potential usefulness, our findings suggest that there are better alternative solutions for effective information recording.

Documentation Burden of Flowsheet Comments

First, although only a small proportion of flowsheet vital sign entries had comments, when considering a median typing speed of 30 words per minute [24], the composition of 700,000 words in 210,000 comments still implies approximately 23,333 minutes (389 hours) of comment documentation. Although, on average, each user spent about 1 minute to document, the time used for writing such comments was substantial for some Epic users because the 5 vital signs were only a small fraction of all the different types of flowsheet entries. This raises the question: Is it necessary to retain such functionality in the EHR system? Some studies have shown that patients who died tended to receive more vital sign comments than other patients [25,26]. However, it is unclear if such an association is useful in practice because it might be the severity of the condition that led to a higher volume of vital sign comments. Alternatively, it is still helpful to simplify the commenting functionality to save the

users' effort in this circumstance. According to our topic analysis, most comments were related to either notification (0.347), the context of performing a measurement (0.307), or notification (0.224). Although context might provide additional information about a measurement, it is unclear whether recording that a health care provider was notified about a measurement contributes to the understanding of a patient's health condition. A deeper review of the identified topics could help determine the need for configuration improvements to the system. For instance, are comments related to notification entered to address potential litigation or is there a true concern for the patient's condition? Are there other signs of patient deterioration? Are comments related to the context of the vital signs already documented elsewhere in a more appropriate location? Are comments related to the inability to take a patient's vital signs fulfilling the nursing mantra of *if it's not documented, it's not done* and a potential legal consequence? Unfortunately, there has been little investigation into the motivation for recording notifications.

Visibility of Medical Information in Comments

In addition, the medical term analysis raises another question: Should such medical information be recorded in flowsheet entry comments? In EHRs, this information should at least be recorded in clinical notes, which ensures that such information could be referred to in the future. However, flowsheet comments are not displayed or automatically pulled into any clinical notes. Rather, the only way to review an existing comment is to locate the flowsheet on the correct date and open the comment display dialog. As such, we suspect that flowsheet comments are seldom reviewed by other health care providers, except the user who made them. Although this needs to be verified (eg, which might be possible through a review of the EHR access logs), it was reported that 5.6% of the alert comments regarding potentially very important clinical safety issues were overlooked [27]. Furthermore, a study showed that only 16% of nursing notes were read by physicians and 38% were read by other nurses [28]. As such, it appears that the flowsheet entry comments might not be in a proper place to store medical-related information. Although only a fraction of notes are examined by others, recording such information in clinical notes seems a better approach to record the information and access it in the future.

Potential Changes to EHR Design

Finally, based on this analysis, we believe there are at least two ways by which this functionality could be better oriented toward a user-centered design. First, the ability to enter free-text comments in a flowsheet row can be removed from the EHR system. Although this may save health care providers' time and effort, it should only be considered after a careful examination of the utility of this functionality, an endeavor that is beyond the scope of this investigation. Second, we suspect that one possible explanation for recording notification is related to the potential for future lawsuits. This notion was highlighted in an interview with 5 acute care nurses, all of whom agreed that notification comments in a flowsheet help to *cover them legally* [21]. If this information must be stored, then it can be designed using structured response fields (eg, in the form of a simple checkbox) such that users do not have to click the comment entry icon, open a dialog box, and then enter comments. This design should be suitable for capturing when a measurement is reliable or unable to obtain as well. Moreover, any medical-related information should be recorded in clinical notes for future reference. We believe that such a design will be much more efficient and ensure that important information can be easily reviewed in the future. However, we acknowledge that a user-centered design approach would help understand the need for, as well as how to improve, the functionality.

Limitation and Future Work

Despite the merits of this work, there are several limitations that we wish to highlight, which could guide future research. First, we only examined the data from a single clinical environment, which may limit the generalizability of our findings. However, this functionality exists across all Epic implementations and is therefore likely to be a widespread phenomenon. Second, we only examined vital sign comments; thus, it is unclear if these findings would hold with other types of flowsheet entries. Third, we only focused on the dominant topic when analyzing the distribution of topic groups within each vital sign. Owing to the brevity of flowsheet comments, topic modeling strategies that are explicitly oriented to handle texts of shorter length should be considered in future investigations. Fourth, we replaced only the misspellings for certain frequent terms. Correcting the spelling errors and resolving aliasing issues (ie, when 2 terms correspond to the same underlying concept) for the entire vocabulary may improve the quality of topic modeling, but determining the best approach to use is beyond the scope of this investigation. Future work may also consider extracting concepts based on specific nursing terminologies in addition to the general medical terms to interpret the comments from a nursing perspective. In addition, it might be beneficial to investigate how the use of comments varies across patient characteristics and settings (eg, comments made during a hospital encounter vs those made outside of a hospital encounter). Moreover, we only examined the comments based on their content. To fully understand this functionality and information, potential future work includes examining the motivation of recording notification in comments, the extent to which such comments would be accessed by other health care providers, and the association between the content of flowsheet comments and patients' health-related behaviors or outcomes.

Conclusions

Documentation burden is a recognizable issue when modern EHR systems are increasingly adopted in health care. One potential solution to reduce such burden is to simplify the existing functionalities of an EHR system. In this study, we examined the nature of vital sign comments in flowsheets using the data generated in the Epic system at VUMC. We found that most of the comments were related to the notification of a result to health care providers, the context of a measurement, and an inability to obtain a vital sign. We also extracted many medical terms (eg, symptoms or medications) from these comments. Considering that flowsheet comments are not displayed or automatically pulled into any clinical notes, we believe that such functionality can be simplified via structured response

fields instead of a text input dialog to reduce health care provider effort.

Acknowledgments

This work was partially sponsored by the National Institutes of Health under grant UL1TR002243.

Authors' Contributions

ZY and PS proposed the research question. AM collected the data. ZY and YL designed and conducted the experiments. ZY drafted the manuscript. ZY, PS, BM, AM, and YL edited and reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental material.
[DOCX File , 68 KB - jmir_v23i3e22806_app1.docx]

Multimedia Appendix 2
Word cloud of medical terms regarding tests. The font size of each word is proportional to its frequency in flowsheet comments.
[PNG File , 70 KB - jmir_v23i3e22806_app2.png]

Multimedia Appendix 3
Word cloud of medical terms regarding problems. The font size of each word is proportional to its frequency in flowsheet comments.
[PNG File , 94 KB - jmir_v23i3e22806_app3.png]

Multimedia Appendix 4
Word cloud of medical terms regarding treatments. The font size of each word is proportional to its frequency in flowsheet comments.
[PNG File , 106 KB - jmir_v23i3e22806_app4.png]

References

1. Health IT Dashboard. URL: https://dashboard.healthit.gov/quickstats/quickstats.php [accessed 2020-05-05]
2. Malte CA, Berger D, Saxon AJ, Hagedorn HJ, Achtmeyer CE, Mariano AJ, et al. Electronic Medical Record Alert Associated With Reduced Opioid and Benzodiazepine Coprescribing in High-risk Veteran Patients. Medical Care 2018;56(2):171-178. [doi: 10.1097/mlr.0000000000000861]
3. CBC. The emerging benefits of electronic medical record use in community-based care. URL: https://www.documentcloud.org/documents/690256-final-infoway-emr-benefits-english-summary.html [accessed 2020-07-05]
4. Huang X, Elston R, Rosa G, Mayer J, Ye Z, Kitchner T, et al. Applying family analyses to electronic health records to facilitate genetic research. Bioinformatics 2018;34(4):635-642. [doi: 10.1093/bioinformatics/btx569]
5. Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan W, Sinsky CA, et al. Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Motion Observations. Ann Fam Med 2017 Sep 11;15(5):419-426. [doi: 10.1370/afm.2121]
6. Read-Brown S, Hribar MR, Reznick LG, Lombardi LH, Parikh M, Chamberlain WD, et al. Time Requirements for Electronic Health Record Use in an Academic Ophthalmology Center. JAMA Ophthalmol 2017 Nov 01;135(11):1250. [doi: 10.1001/jamaophthalmol.2017.4187]
7. Collins S, Couture B, Kang MJ, Dykes P, Schnock K, Knaplund C, et al. Quantifying and visualizing nursing flowsheet documentation burden in acute and critical care. 2018 Presented at: AMIA Annu Symp Proc; 2018; Washington D.C p. 348-357.
8. Patients Not Paperwork ? Bureaucracy affecting nurses in the NHS. 8 Cunningham L, Kennedy J, Nwolisa F, Wike C. URL: https://www.qualitasconsortium.com/index.cfm/reference-material/fundamentals/patients-not-paperwork/ [accessed 2020-07-05]
9. Padden J. Documentation Burden and Cognitive Burden. CIN: Computers, Informatics, Nursing 2019;37(2):60-61. [doi: 10.1097/cin.0000000000000522]
10. Strategy on Reducing Burden Relating to the Use of Health IT and EHRs. HealthIT.gov. URL: https://www.healthit.gov/topic/usability-and-provider-burden/strategy-reducing-burden-relating-use-health-it-and-ehrs

11. Fratzke J, Tucker S, Shedenhelm H, Arnold J, Belda T, Petera M. Enhancing Nursing Practice by Utilizing Voice Recognition for Direct Documentation. JONA: The Journal of Nursing Administration 2014;44(2):79-86. [doi: 10.1097/nna.0000000000000030]
12. Awan SK, Dunoyer EJ, Genuario KE, Levy AC, O'Connor KP, Serhatli S, et al. Using voice recognition enabled smartwatches to improve nurse documentation. 2018 Presented at: InSystemsInformation Engineering Design Symposium (SIEDS); 2018; Charlottesville, VA p. 159-164. [doi: 10.1109/sieds.2018.8374728]
13. Warner JL, Smith J, Wright A. It's Time to Wikify Clinical Documentation. Academic Medicine 2019;94(5):645-650. [doi: 10.1097/acm.0000000000002613]
14. Luh JY, Thompson RF, Lin S. Clinical Documentation and Patient Care Using Artificial Intelligence in Radiation Oncology. Journal of the American College of Radiology 2019 Sep;16(9):1343-1346. [doi: 10.1016/j.jacr.2019.05.044]
15. Fabbri D, Ehrenfeld JM. Hands Free Automatic Clinical Care Documentation: Opportunities for Motion Sensors and Cameras. J Med Syst 2016 Aug 16;40(10):213. [doi: 10.1007/s10916-016-0570-x]
16. Bates DW, Landman AB. Use of Medical Scribes to Reduce Documentation Burden. JAMA Intern Med 2018 Nov 01;178(11):1472. [doi: 10.1001/jamainternmed.2018.3945]
17. Tran BD, Chen Y, Liu S, Zheng K. How does medical scribes? work inform development of speech-based clinical documentation technologies? A systematic review. J Am Med Inform Assoc in press 2020;27(5):808-817. [doi: 10.1093/jamia/ocaa020]
18. Shah T, Patel-Teague S, Kroupa L, Meyer AND, Singh H. Impact of a national QI programme on reducing electronic health record notifications to clinicians. BMJ Qual Saf 2018 Mar 05;28(1):10-14. [doi: 10.1136/bmjqs-2017-007447]
19. Healthcare IT News. 19 How one hospital tweaks its EHR to fight alert fatigue. 2015. URL: https://www.healthcareitnews.com/news/how-one-hospital-tweaks-its-ehr-fight-alert-fatigue [accessed 2020-07-05]
20. Winden TJ, Chen ES, Monsen KA, Wang Y, Melton GB. Evaluation of Flowsheet Documentation in the Electronic Health Record for Residence, Living Situation, and Living Conditions. In: AMIA Jt Summits Transl Sci Proc. 2018 Presented at: p. 236-245 URL: http://europepmc.org/abstract/MED/29888079
21. Collins SA, Fred M, Wilcox L, Vawdrey DK. Workarounds used by nurses to overcome design constraints of electronic health records. In: The 11th Int Congr Nurs Inform. 2012 Presented at: The 11th Int Congr Nurs Inform; 2012; Montreal Canada p. 93.
22. Yin Z, Song L, Clayton EW, Malin BA. Health and kinship matter: Learning about direct-to-consumer genetic testing user experiences via online discussions. PLoS ONE 2020 Sep 8;15(9):e0238644. [doi: 10.1371/journal.pone.0238644]
23. Soysal E, Wang J, Jiang M, Wu Y, Pakhomov S, Liu H, et al. CLAMP - a toolkit for efficiently building customized clinical natural language processing pipelines. J Am Med Inform Assoc 2018 Mar 01;25(3):331-336 [FREE Full text] [doi: 10.1093/jamia/ocx132] [Medline: 29186491]
24. LiveChat. 24 Typing Speed Test - Check your typing skills!. 2016. URL: https://www.livechatinc.com/typing-speed-test/ [accessed 2020-07-05]
25. Collins SA, Cato K, Albers D, Scott K, Stetson PD, Bakken S, et al. Relationship Between Nursing Documentation and Patients' Mortality. American Journal of Critical Care 2013 Jul 01;22(4):306-313. [doi: 10.4037/ajcc2013426]
26. Collins SA, Vawdrey DK. "Reading between the lines" of flow sheet data: nurses' optional documentation associated with cardiac arrest outcomes. Applied Nursing Research 2012 Nov;25(4):251-257. [doi: 10.1016/j.apnr.2011.06.002]
27. Chused AE, Kuperman GJ, Stetson PD. Alert override reasons: a failure to communicate. : AMIA Annu Symp Proc. 2008; 2008 Presented at: AMIA Annu Symp Proc; 2018; Washington, DC p. 111-115.
28. Hripcsak G, Vawdrey DK, Fred MR, Bostwick SB. Use of electronic clinical documentation: time spent and team interactions. Journal of the American Medical Informatics Association 2011 Mar 01;18(2):112-117. [doi: 10.1136/jamia.2010.008441]

Abbreviations

BP: blood pressure
CLAMP: Clinical Language Annotation, Modeling, and Processing
EHR: electronic health record
LDA: latent Dirichlet allocation
Pulse: pulse rate
Resp: respiration rate
RQ: research question
SpO_2: oxygen saturation
Temp: body temperature
VUMC: Vanderbilt University Medical Center

Edited by G Eysenbach, R Kukafka; submitted 23.07.20; peer-reviewed by S Rossetti, M Claes; comments to author 21.08.20; revised version received 11.10.20; accepted 18.01.21; published 04.03.21.

Please cite as:
Yin Z, Liu Y, McCoy AB, Malin BA, Sengstack PR
Contribution of Free-Text Comments to the Burden of Documentation: Assessment and Analysis of Vital Sign Comments in Flowsheets
J Med Internet Res 2021;23(3):e22806
URL: https://www.jmir.org/2021/3/e22806
doi:10.2196/22806
PMID:33661128

Review

eHealth Applications to Support Independent Living of Older Persons: Scoping Review of Costs and Benefits Identified in Economic Evaluations

Sandra Sülz[1], PhD; Hilco J van Elten[1], PhD; Marjan Askari[1], PhD; Anne Marie Weggelaar-Jansen[1,2], PhD; Robbert Huijsman[1,3], PhD

[1]Erasmus School of Health Policy & Management, Rotterdam, Netherlands

[2]Clinical Informatics, Eindhoven University of Technology, Eindhoven, Netherlands

[3]Geriant, Heerhugowaard, Netherlands

Corresponding Author:
Sandra Sülz, PhD
Erasmus School of Health Policy & Management
PO Box 1738
Rotterdam, 3000 DR
Netherlands
Phone: 31 104088531
Email: sulz@eshpm.eur.nl

Abstract

Background: eHealth applications are constantly increasing and are frequently considered to constitute a promising strategy for cost containment in health care, particularly if the applications aim to support older persons. Older persons are, however, not the only major eHealth stakeholder. eHealth suppliers, caregivers, funding bodies, and health authorities are also likely to attribute value to eHealth applications, but they can differ in their value attribution because they are affected differently by eHealth costs and benefits. Therefore, any assessment of the value of eHealth applications requires the consideration of multiple stakeholders in a holistic and integrated manner. Such a holistic and reliable value assessment requires a profound understanding of the application's costs and benefits. The first step in measuring costs and benefits is identifying the relevant costs and benefit categories that the eHealth application affects.

Objective: The aim of this study is to support the conceptual phase of an economic evaluation by providing an overview of the relevant direct and indirect costs and benefits incorporated in economic evaluations so far.

Methods: We conducted a systematic literature search covering papers published until December 2019 by using the Embase, Medline Ovid, Web of Science, and CINAHL EBSCOhost databases. We included papers on eHealth applications with web-based contact possibilities between clients and health care providers (mobile health apps) and applications for self-management, telehomecare, telemedicine, telemonitoring, telerehabilitation, and active healthy aging technologies for older persons. We included studies that focused on any type of economic evaluation, including costs and benefit measures.

Results: We identified 55 papers with economic evaluations. These studies considered a range of different types of costs and benefits. Costs pertained to implementation activities and operational activities related to eHealth applications. Benefits (or consequences) could be categorized according to stakeholder groups, that is, older persons, caregivers, and health care providers. These benefits can further be divided into stakeholder-specific outcomes and resource usage. Some cost and benefit types have received more attention than others. For instance, patient outcomes have been predominantly captured via quality-of-life considerations and various types of physical health status indicators. From the perspective of resource usage, a strong emphasis has been placed on home care visits and hospital usage.

Conclusions: Economic evaluations of eHealth applications are gaining momentum, and studies have shown considerable variation regarding the costs and benefits that they include. We contribute to the body of literature by providing a detailed and up-to-date framework of cost and benefit categories that any interested stakeholder can use as a starting point to conduct an economic evaluation in the context of independent living of older persons.

(J Med Internet Res 2021;23(3):e24363) doi:10.2196/24363

KEYWORDS

cost; benefit; eHealth; aged; economic evaluation

Introduction

Background and Motivation

The use of information and communication technologies in health care is regarded as an important piece of the puzzle of increasing health care costs and demand [1], particularly if it targets older persons with substantial health care costs [2]. Induced by an aging population that stays home longer, an increase in self-management, and a changing role of informal caregivers, the demand for health care delivery, and the role of technology are changing rapidly. To contain health care costs and maintain the quality of care and living, governments direct policies to stimulate eHealth to increase and support self-management [3]. In this study, eHealth is defined in line with the description by Eysenbach [4] and the different taxonomies described by Oh et al [5]: eHealth is at the intersection of medical informatics, public health, and business and offers health services to support care delivery, manage care, promote prevention, and educate; it is delivered or enhanced through the internet and related technologies (eg, domotics, wearables, and sensors). In the domain of older persons living at home, we define eHealth as web-based contact possibilities between the clients and health care providers and applications for self-management, telehomecare, telemedicine, telemonitoring, and telerehabilitation.

eHealth has shown to be valuable in promoting medication adherence and improving self-management in the population of older persons [6]. In addition, eHealth can be used to monitor clinical signs, collect health information, support users in activities related to their health, and promote a healthy lifestyle or arrange remote consultations [7-10]. Growing internet access, increasing use of mobile apps, and current technology trends create opportunities for novel services and new forms of health care through eHealth [11,12]. Governments are also increasingly funding initiatives that replace traditional care with alternatives that use information and communication technologies to remotely monitor and deliver health care services. Primary funding motivation is economic in nature—promoting preventive measures to avoid costly consequences and stimulate efforts to increase access to care [13].

Although eHealth is frequently considered a promising development, these applications are not without considerable costs. eHealth equipment must be purchased, and systems must be operated and maintained. Data recorded by eHealth should be monitored. However, frequently, the stakeholder who benefits from the application is not the same stakeholder who is paying for it. Costs and benefits affect different stakeholders and potentially also at different points in time; therefore, the economic interests of stakeholders are often not aligned. From the health provider perspective, such an investment does not make an economic sense, whereas it might be highly valuable from a societal perspective, considering the total benefits and costs regardless of where they occur. Any assessment of the value of an eHealth application, therefore, requires considering multiple stakeholders in a holistic and integrated assessment of all costs (ie, the direct and indirect and short- and long-term costs) and all benefits (ie, the direct and indirect and short- and long-term *gains*) of an eHealth application. This rationale represents the core of economic evaluations in health care [14]. Economic evaluations in health care come in different forms such as cost-effectiveness analyses (CEAs), cost utility analyses (CUAs), and cost-benefit analyses (CBAs). Regardless of the type, they have in common considering both costs and benefits, that is, what we have to give up *and* what we will gain. The main difference is the way in which *gain* is incorporated: CEAs consider a one-dimensional measure of the *gain*, which only allows for a comparison of programs with the same effect measures. CUAs assess gain through utility, frequently in quality-adjusted life years, which is comparable between health programs. CBAs assess the gain monetarily, that is, costs and outcomes are directly on the same scale [14,15].

Research Objective

If an interested stakeholder wants to conduct an economic evaluation, that is, apply such an evaluation to a specific case, he or she must first identify the relevant cost and benefit elements. For this identification step, the body of literature can serve as a valuable source of information. Currently, it is unclear what types of information are available in the literature about the relevant cost and benefit elements of eHealth. Therefore, we conducted a scoping review to systematically map the literature in this area and to develop an up-to-date framework for conducting all-inclusive economic evaluations of eHealth applications that support independent living of older persons.

Methods

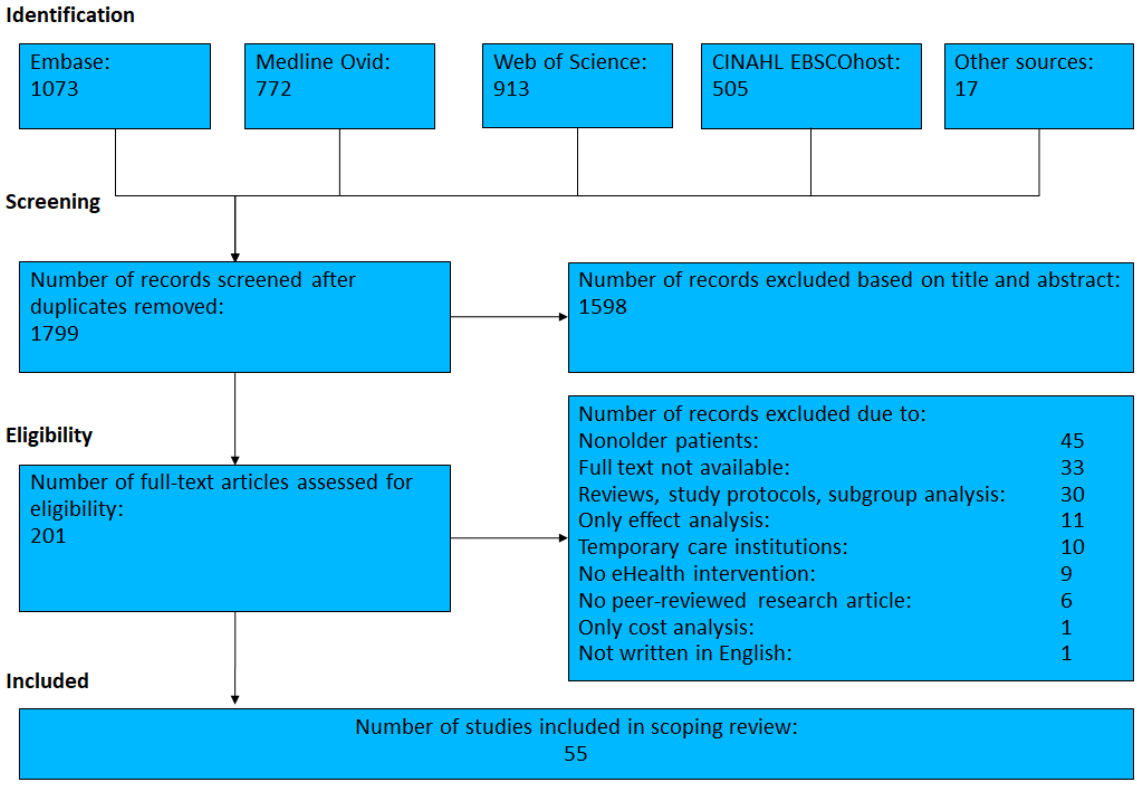
Scoping Review

Our research methodology was drafted using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis) protocols, specifically for scoping reviews. Among other things, scoping reviews are appropriate for identifying key characteristics and factors related to a concept [16]. Consequently, they lend themselves naturally to our research objectives.

Search Query and Inclusion and Exclusion Criteria

For this study, we identified papers published until December 2019 using the Embase, Medline Ovid, Web of Science, and CINAHL EBSCOhost databases. Additional papers were identified by scanning the references of the identified papers (indicated as other sources in the PRISMA flowchart in Figure 1). An experienced librarian developed search strings with some unique features to combine search terms effectively [17,18] and conducted the search. Our search terms are derived from the inclusion and exclusion criteria specified later, and the complete search query is provided in Multimedia Appendix 1.

Figure 1. Flowchart of paper selection.



The inclusion and exclusion criteria were specified to guide the identification process. We included original papers that were peer reviewed, had an empirical or prescriptive nature, and were written in English. Research protocols, commentaries, and editorial papers were excluded from the study. We also excluded reviews to avoid duplicate findings and avoid relying on the review's interpretation of the cost and benefit labels. Concerning the patient population, we included papers that focused on patients with an average age of at least 65 years and who were living independently at their usual place of residence. Studies were excluded if they considered nonolder persons (ie, infants, adolescents, and a sample mean age younger than 65 years) or if the patient population received care in any institutionalized form (ie, hospitals, nursing homes, rehabilitation clinics, and hospice). In terms of the eHealth application, we focused on eHealth applications with web-based contact possibilities between clients and health care providers, including mobile health (mHealth) apps, and applications for self-management, telehomecare, telemedicine, telemonitoring, telerehabilitation, and active healthy aging technologies for older persons. Papers with the following technologies were excluded: papers describing technology not connected to the internet (eg, implantable cardioverter defibrillator) and papers on health information systems, electronic health records, robotics, and telephone consults only. To obtain a comprehensive overview, we did not use a strict definition of economic evaluation and did not constrain our search to particular types of economic evaluations. The only constraint on which we relied was that economic evaluations required both costs and benefits. Therefore, we included papers that measured at least one monetary aspect and one benefit of eHealth applications in any shape or form (eg, medical expenditures and analyzing clinical effects). Papers that only described or analyzed costs at an aggregated level or (clinical) effects were excluded.

Screening and Eligibility

Initially, 2 of the authors (SS and HE) screened a small sample simultaneously to align the assessment. Subsequently, SS and HE conducted abstract and title screening, with each screening half of the identified references. The studies were labeled *not relevant* or *potentially relevant* based on the inclusion criteria. A random sample of the references was blindly double-checked by AW and RH. The random set has been created by selecting every 10th article from the set of annually ordered articles. Interrater reliability was evaluated using the Cohen kappa (κ) index, which is a robust statistic useful for interrater reliability testing [19]. This double screening yielded a *fair* overlap between the assessments by SS or HE and AW or RB (Cohen κ index=0.28). After realigning the inclusion and exclusion criteria and rescreening all of the references with previous disagreements (conducted by AW), substantial improvements in the alignment were achieved (Cohen κ index=0.75).

In the next screening round, SS and HE conducted the screening based on full texts, and both authors went independently through all of the references initially labeled as *potentially relevant*. This double-blind screening yielded a *substantial* overlap (Cohen κ index=0.64). SS and HE discussed all of the references for which their assessments diverged and thus resolved disagreements.

Data Extraction and Categorization

Two authors (SS and HE) extracted the following information from the studies: patient population, country, type of eHealth intervention, type of analysis that the papers report performing, the specific cost types that the studies considered, and the

specific benefits, gains, and consequences that the studies included. Subsequently, we categorized the costs and benefits. We started with 4 categories pertaining to (1) implementation activities for the eHealth application, (2) operating and maintenance activities for the eHealth application, (3) processes of health care delivery, and (4) outcomes. While extracting the information, we further refined these categories depending on activities, stakeholders, and institutions such that the subcategories were mutually exclusive but sufficiently broad to capture related items.

Results

Study Selection

The search and screening processes are shown in Figure 1. The database and manual searches resulted in 1799 papers. After the first round of removal of duplicate papers and title and abstract screening, the remaining 201 papers were subjected to full-text reading. After the second step, 55 papers met our inclusion criteria and were included in the final set for analysis. The included references were published in the 2000-2019 time frame, with 5 included papers from the first 5-year period (2000-2004), 12 between 2005 and 2009, 17 between 2010 and 2014, and 21 between 2015 and 2019.

Patient Characteristics

The patient characteristics were rather broad; however, most studies focused on chronic or cardiovascular diseases. Of the 55 references, 13 had a patient population with various chronic health problems [13,20-31]. Cardiovascular problems received specific attention in 17 studies [32-48]. In 10 studies, the focus was on chronic obstructive pulmonary disease [49-58], and in 2 studies, the focus was on chronic skin problems [59,60]. Other studies considered diabetes [61], age-related macular degeneration [62], post-knee arthroplasty patients [63], Parkinson disease [64], and terminal patients [65]. Five studies did not specify any health conditions or diseases [66-70]. Of the 55 studies, 3 focused on mental and behavioral disorders such as anxiety [71], dementia [72], and depression [73].

Country

The studies in English were geographically clustered, with 26 studies conducted in North America, among which 20 occurred in the United States [20-23,25-28,30,31,33,34,37,38,42,61,62,64,67,68] and 6 in Canada [29,40,49,51,63,66]. Among the 22 studies conducted in Europe, they were spread across countries, with 5 studies conducted in England and the United Kingdom [24,41,43,44,53]; 3 each in Denmark [50,52,58], Italy [32,45,55], the Netherlands [35,39,72]; 2 each in Austria [46,60] and Germany [47,54]; and 1 each in France [59], Spain (although not explicitly stated in the paper) [57], Norway [70], and Sweden [69]. Finally, 4 studies took place in Australia [36,56,71,73] and 3 in Asia, among which 2 were in Japan [13,65] and 1 in Taiwan [48]. With the study in Taiwan being the only one conducted in a country that is not a part of the Organization for Economic Cooperation and Development, there was a strong emphasis on economically strong countries with aging populations.

Type of eHealth Interventions

The majority of included studies focused on telemonitoring or remote monitoring involving the measurement of vital statistics and the transmission of patient data followed by an assessment—either automatically or manually—and triggering action by health care professionals if required [22-25,27-30,32-35,38-43,45-51,53,54,56,57,62,65,68,70]. In addition, other eHealth forms included in our review related to video consultations and virtual visits [20,21,23,30,37,42,48,55,59,61,63,64,69]; deployment of sensor technology to analyze behavioral patterns and wireless transmitters [13,26,48,61,66,67,70,72]; email messaging services and web portal access [44,69,71,73]; online disease management courses or resources [44,71,73]; internet-delivered cognitive behavioral therapy [71,73]; remotely supervised rehabilitation activities, such as assistant mHealth [36,52,58]; and digital data transmission [60]. Note that some studies blended various eHealth forms and applications and that there was some ambiguity in the terminology, with telehealth often being used interchangeably with telemedicine, telemonitoring, or remote monitoring. The specific eHealth interventions analyzed in the studies are described in Multimedia Appendix 2 [13,20-73].

Type of Analysis

The included studies indicated various types of analysis. In total, 18 studies reported CEAs [24,25,27,34,35,39-41,44,46,50,53,54,59,60,70,71,73], 2 studies combined CEAs with clinical or budget impact analyses [48,62], 2 studies reported CUAs [36,52], and 1 study reported a CBA [13]; 8 studies denoted the evaluation as cost analysis [26,45,61,63,67-69,72], the term *cost minimization* was used in 4 studies [29,49,51,65], and 1 study reported relying on cost consequence analysis [43]. The remaining 17 studies stated that they evaluated a range of cost and outcome measures [21,23,28,30-33,37,38,42,47,55-58,64,66], and 2 studies relied on case studies to outline benefits, saving, expenditures, and outcomes [20,22].

Of the 55 included studies, 8 relied on Markov modeling and simulation models with a time frame of 1 year [40,65], 5 years [13,34], 10 years [43,62], 20 years [35], up to 30 years [41]. Only a few studies without Markov modeling explicitly stated the time frame of the economic evaluation, with 1 study capturing a period of multiple weeks [63] and 6 studies with a 1-year time frame [24,39,44,45,50,54]. The remaining 40 studies did not indicate the time frame explicitly, and the period of data collection was, if applicable, used as a proxy for the time frame: 1-6 months [22,25,36-38,42,51,56,57,60,66], 7-12 months [28,29,48,52,58,59,64,67-69,71-73], 13-18 months [20,21,31,53], 19-24 months [27,49,61], and beyond 24 months [26,30,32,33,46,47,55,70]. In 1 study, the period of data collection was insufficiently described [23].

Specific Cost and Benefit Types

When we consider eHealth applications, what direct and indirect costs and benefits (or consequences) might be relevant to consider? Tables 1-3 provide an answer by outlining the different cost and benefit types that the studies included in our review considered. Table 1 shows the eHealth intervention costs categorized into implementation and operating activities. Table

2 provides an overview of the consequences of eHealth and focuses on resource usage. Finally, Table 3 depicts eHealth consequences in terms of outcomes categorized by stakeholder group.

Table 1. Intervention costs of eHealth applications.

Intervention costs	Considered by number of studies	References
Implementation activities		
Device purchase (monitoring equipment, videoconferencing equipment, etc)	32	[13,22-25,29-31,35-37,39-41,44,47,50-53,55,56,59,61,63,65,66,69-73]
License or software or initial fee purchase	7	[13,24,29,50,51,59,72]
Equipment installation	18	[13,24,29,34,35,38,50-53,55,61,63,65,66,69,70,72]
Training or education of operators	10	[24,25,29,35,44,50,52,53,61,70]
Technician travel time to install equipment	5	[30,36,40,52,63]
Operating activities		
Maintenance (server, host, call center, or station)	18	[13,24,26,28,29,35,37,39,40,43,50,51,53,55,59,65,66,68,69]
Periodic fees (for licenses, insurance, etc)	15	[24,29,31,35,37,38,43,50,55,59,61,68,71-73]
Medical staffing: reviewing or assessing or intervening	26	[13,24,29,32,34-36,38,40,41,44,50,53,55-57,61-63,65,66,69-73]
Nonmedical staffing: technical support	14	[13,24,29,39,40,50,51,53,61,63,66,68,69,72]
Technician travel time to maintain equipment	1	[24]

Table 2. Intervention consequences—resource usage.

Intervention consequences	Considered by the included studies, n	References
eHealth usage by patient		
Televisits: number or duration of visits	17	[20,22,23,25,30-32,35,37,39,44,45,51,56,61,64,70]
Health resource usage by patient—community health services or primary care		
Travel time or transportation costs (eg, for ambulance)	9	[39,44,46,53,55,59,60,64,67]
General practitioner: number or duration of visits	16	[24,28,31,35,39,40,44,50,52,55-57,59,68,71,73]
Walk-in center: number or duration of visits	2	[24,44]
Physiotherapist: number of sessions	5	[24,39,52,59,63]
Psychologist: number of sessions	2	[24,39]
Community nurse: number or duration of visits	4	[24,39,44,50]
Home care: number or duration of visits	27	[21-25,28-32,34,37-39,49-51,55-57,61,64-67,69,72]
Meals on wheels	1	[24]
Day services		
Day care	2	[24,69]
Same-day surgeries	1	[67]
Institutionalized care		
Rehabilitation clinics: number or duration of admissions	2	[50,54]
Skilled nursing facilities: number or duration of admissions	6	[27,62,67,69,70,72]
Long-term care: number or duration of admission	2	[24,67]
Hospice: number or duration of admissions	2	[21,35]
Hospital use		
Emergency department: number of visits	21	[21,24,27-29,31,33,35,37-39,42,49,50,52,53,55-57,67,68]
Outpatient clinic: number or duration of visits to specialists	21	[21,24,27,28,34,35,40,43,45,46,48,50,52-54,56,59,60,62,67,68]
Hospital: number or duration of admissions	40	[21,24,25,27-31,33,34,36-59,65-68,70,71,73]
Intensive care unit: Admissions	2	[55,57]
Drug treatment and laboratory diagnostics		
Medication, prescriptions, or medical supplies	16	[21,24,31,38,39,43,44,50,52,54,55,59,62,67,71,73]
Laboratory	3	[66,67]

Table 3. Intervention consequences—stakeholder outcomes.

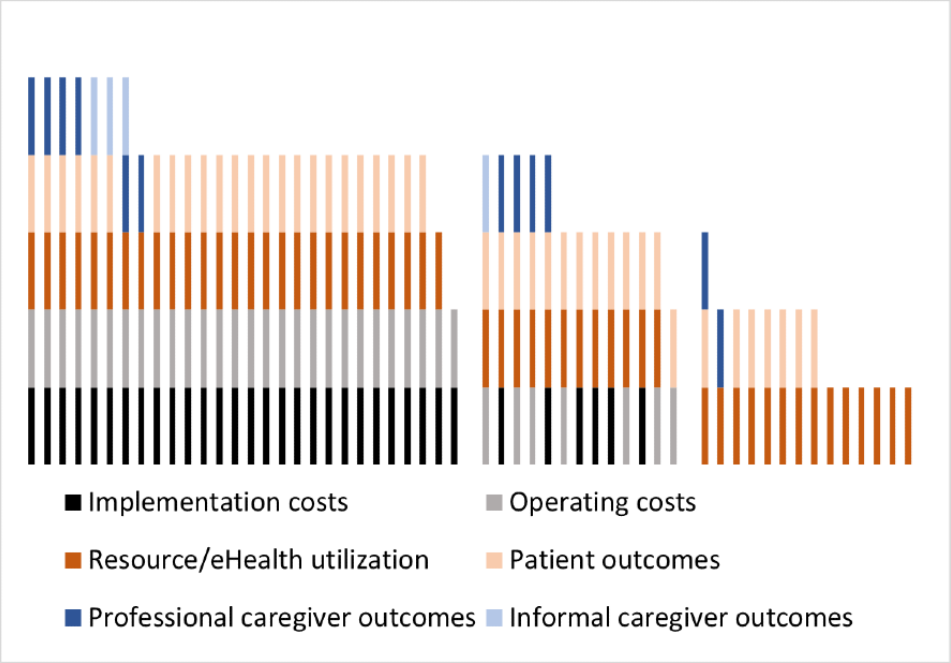
Intervention consequences	Considered by the included studies, n	References
Patient or client outcomes		
Physical health status (mortality, morbidity, cardiovascular events, exacerbations, etc)	14	[30,32-34,43,44,46,47,54,55,59-61,68]
Psychological health status (anxiety, depression, or empowerment)	4	[22-24,66]
QALYs[a]	12	[24,35,36,39-41,44,50,52,62,71,73]
Quality of life (if not measured in QALYs but differently)	11	[24,28,33,37,38,42,47,56,61,64,66]
Setting-specific quality of care indicators	1	[31]
Satisfaction (with the device or eHealth service)	8	[20,23,25,29,31,40,51,66]
Satisfaction (in general)	5	[22,28,30,37,61]
Patient experience or perceived benefits	3	[57,65,69]
Well-being	1	[72]
Time spent in the usual place of residence	1	[26]
Transfer to a different level of care	1	[30]
Time absent from work (productivity loss or loss of income)	2	[44,62]
Device-related technical events	2	[43,47]
Professional caregivers		
Satisfaction with the device	2	[20,25]
Satisfaction in general	1	[30]
Travel time to patient's home	11	[20,25,28,30-32,49,51,56,63,65,70]
Informal caregivers		
Time absent from work (productivity loss)	2	[62,70]
Burden	1	[66]
Well-being	1	[72]
Transfer payments		
Attendance allowance (recipients receive payments to manage their own health)	1	[69]
Respite care (payments made to relieve informal caregivers from providing care)	1	[69]

[a]QALYs: quality-adjusted life years.

Level of Integration

Although Tables 1-3 provide an extensive overview of cost and consequence aspects already considered in previous studies, they indicate that there is diversity in how many different aspects the studies have considered. For all the studies, we determined whether they included at least one element of implementation costs, operating costs, and consequences such as eHealth and resource usage, patient outcomes, professional caregiver outcomes, and informal caregiver outcomes. The studies had different foci and different levels of integration, as outlined by the diversity in Figure 2. Each study is represented by 1 bar decomposed into the cost and consequence subcategories considered in the study. The group of 28 studies on the left considered at least one implementation and one operating activity, in addition to the consequences of the eHealth intervention [13,24,29,31,34-41,44,50,51,53,55,56,59,61,63,65,66,69-73]. The middle group of 13 studies contains either implementation or operating activities, in addition to consequences [22,23,25,26,28,30,32,43,47,52,57,62,68]. Finally, the 14 studies on the right do not consider implementation and operating activities but only focus on the consequences [20,21,27,33,45,46,48,49,54,58,60,64,67].

Figure 2. Level of integration in 55 included studies (represented as bars).



Prominent and Lacking Features

Tables 1-3 also show that some components have received more attention than others. Concentrating on the different consequences, we observed the following: If patient outcomes are considered, the focus lies on physical outcomes [30,32-34,43,44,46,47,54,55,59-61,68] and quality of life or quality-adjusted life years [24,35,36,39-41,44,50,52,62,71,73]. Productivity loss because of time absent from work received less attention, with only 4% (2/55) of the studies incorporating it [44,62]; however, this outcome is understandable given that our review focused on the population of older persons, which predominantly no longer participates in the labor market. In terms of resource utilization, there is a strong emphasis on home care visits, indicated by 49% (27/55) of the included studies [21-25,28-32,34,37-39,49-51,55-57,61,64-67,69,72], and hospital usage covered, indicated by 73% (40/55) of the included studies [21,24,25,27-31,33,34,36-59,65-68,70,71,73]. This outcome can be explained by many of the included studies relying on remote monitoring or virtual visits to substitute for home care visits or to prevent exacerbations that lead to hospital admissions.

Comparing the consequences for professional caregivers and informal caregivers, it becomes obvious that certain elements are missing across all included studies. For instance, 4% (2/55) of the included studies considered the satisfaction of professional caregivers (in general or with the device) [20,25]; however, the satisfaction of informal caregivers was not captured in any of the included studies. Similarly, for professional caregivers, 20% (11/55) of the included studies captured the travel time to patients' homes [20,25,28,30-32,49,51,56,63,65,70]; however, this aspect was neglected for informal caregivers. Conversely, for informal caregivers, emotional burden and well-being were captured by 4% (2/55) of the included studies [66,72], whereas none of the included studies focused on these 2 aspects for professional caregivers.

Our final observation relates to the research and development (R&D) costs of eHealth applications. In fact, none of the included studies considered R&D costs, which is understandable from the perspective that these costs have already been spent by the time that the eHealth application is set up and running. However, it implicitly assumes that R&D costs only occur before the eHealth intervention is set up and that there is no ongoing refinement.

Discussion

Principal Findings

We conducted this review to establish a framework of costs and benefits considered in the economic evaluations of eHealth applications that support the independent living of older persons. Our search identified 55 papers that conducted economic evaluations. All of the identified papers focused on independent living of older persons in their role as patients with one or more chronic conditions. The identified papers considered a range of different types of costs and benefits. Costs pertain to implementation activities and operating activities related to eHealth applications. Benefits (or consequences) can be categorized according to stakeholder groups, that is, patients, caregivers, and health care provider organizations. These benefits can be further divided into stakeholder-specific outcomes and resource utilization. Some cost and benefit types have received more attention than others. For instance, patient outcomes are predominantly captured via quality-of-life considerations and various types of physical health status indicators. From a resource utilization perspective, a strong emphasis is placed on home care visits and hospital usage. One reason for this emphasis is the frequency in which studies focus on remote monitoring to prevent unnecessary hospital admissions or to substitute for home care visits.

Our data extraction also revealed a set of elements that have not been considered across all of the identified papers, including

travel time and satisfaction of informal caregivers, emotional burden and well-being of professional caregivers, and last but not least, the R&D costs of the eHealth application. The reasons why these aspects have been neglected can be manifold: they might have been irrelevant from the perspective from which the economic evaluation was carried out, they might have been unobservable because of long time lags beyond the time frame of the evaluation, or it might have been infeasible to capture and quantify these aspects because of methodological obstacles or data unavailability. That the initial R&D costs of the eHealth applications have not been considered can be justified with the sunk cost argument—by the time that the evaluation takes place, the R&D costs have already been spent. In this sense, the eHealth application is considered a static technology. In the longer run, however, the eHealth application might require an upgrade to comply with new laws and regulations, for reasons pertaining to data privacy, data security, data storage, or improvements in usability. Whether these upgrade costs are indeed relevant for future economic evaluations depends on the perspective and time frame.

Like any other technology, eHealth applications are subject to change. Fueled by the increasing availability of data, changes in law and regulations and improved usability, new fields, and areas of applications might emerge. Our framework is based on eHealth applications currently in place. Future eHealth applications might generate costs and benefits that are different from the eHealth applications on which our framework is based.

Implications for Research and Practice

Economic evaluations of eHealth applications are gaining momentum, as indicated by the increasing number of publications and reviews [74,75]. Health economic frameworks and principles are described, and the steps to measure costs and benefits are emphasized [14,15]. Our review directly connects to the measurement aspect by focusing on its first step, that is, the identification of costs and benefits. We contribute to the body of literature by providing a detailed and up-to-date framework of cost and benefit categories.

If we consider eHealth, there are many stakeholders involved, such as patients, eHealth suppliers (formal and informal) caregivers, funding bodies, health authorities, and so on. Notably, these stakeholders are likely to attribute different values to eHealth applications because they are affected differently by their costs and benefits. This fact has consequences for investment and funding decisions, and it has long been argued that decision making remains hampered by the lack of reliable cost and benefit estimations [14,15]. To obtain reliable cost and benefit estimates, our framework could be of help by providing a starting point for the identification process. Our framework can support this goal; however, the measuring and valuing process must be context-specific and tailored to the case at hand. How to value components that, for instance, do not lend themselves naturally to being quantified in countable units, such as the feeling of safety [76], is worthy of study in itself and beyond the scope of our review. The variety that we observe in terms of costs and consequences considered in the economic evaluations and the type of analysis performed and the extent to which the costs and consequences are integrated indicates that the field has not yet reached consensus on a standard procedure for determining eHealth value. Whether and to what extent the observed variety is actually linked to the quality of the economic evaluation is an interesting avenue for future research.

Financial costs directed to patients or informal caregivers might be an obstacle to using the eHealth service or application or continuation of usage of these services, especially in countries where mandatory health insurance coverage is lacking. Although a recent study in the Netherlands identified finance as a factor not significantly related to intention to use medical apps among older persons [77], future studies should focus on how the costs are covered and who should pay for the direct or indirect costs of eHealth.

Conclusions

A holistic and reliable value assessment of eHealth applications requires a profound understanding of the applications' costs and benefits. The first step in measuring costs and benefits is identifying the relevant costs and benefit categories that the eHealth application affects. We conduct a scoping review to support this identification process by providing an overview of the direct and indirect costs and benefits that economic evaluations have incorporated so far. Our cost-and-benefit framework is particularly useful in the context of eHealth applications that support the independent living of older persons. As this patient group is expected to be increasingly targeted to contain health care costs, expanding the scope of eHealth applications to broader populations, rather than only diagnosed patient groups, and assessing the value of eHealth technologies are of the utmost importance for making well-informed investment and funding decisions.

Acknowledgments

This work was supported by Dutch ZonMw grant number 639003401. The authors would like to thank Wichor Bramer from the Erasmus MC Medical Library for developing and updating the search strategies. The authors further benefited from discussions with the project team Value-Driven eHealth, that is, Rachita Auragh, Karlijn Cranen, Tyrell Creton, Marleen de Mul, and Mildred Visser and their collaboration with Henk Herman Nap (Vilans) and Eveline Wouters (Tilburg University and Tranzo).

Authors' Contributions

All authors have contributed to the conception and design of the review, that is, specifying the inclusion and exclusion criteria (SS, HE, MA, AW, and RH), abstract and title screening (SS, HE, AW, and RH), full-text screening, and data extraction (SS and

HE). SS and HE drafted the manuscript. MA, AW, and RH worked on the revisions of the paper. All authors provided feedback on the multiple draft versions. SS finalized the revisions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 114 KB - jmir_v23i3e24363_app1.pdf]

Multimedia Appendix 2
Type of eHealth intervention, patient population, country, and analysis type.
[PDF File (Adobe PDF File), 175 KB - jmir_v23i3e24363_app2.pdf]

References

1. Granja C, Janssen W, Johansen MA. Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature. J Med Internet Res 2018 May 01;20(5):e10235. [doi: 10.2196/10235]
2. Wouters M, Swinkels I, van Lettow B, de Jong J, Sinnige J, Brabers A, et al. E-health in verschillende snelheden Health-monitor. 2018. URL: https://www.nivel.nl/nl/publicatie/e-health-verschillende-snelheden-ehealth-monitor-2018 [accessed 2021-02-19]
3. van Gemert-Pijnen L, Kelders SM, Kip H, Sanderman R. eHealth Research, Theory and Development A Multi-Disciplinary Approach. London: Routledge; 2018.
4. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun 18;3(2):e20. [doi: 10.2196/jmir.3.2.e20]
5. Oh H, Rizo C, Enkin M, Jadad A, Powell J, Pagliari C. What Is eHealth (3): A Systematic Review of Published Definitions. J Med Internet Res 2005 Feb 24;7(1):e110. [doi: 10.2196/jmir.7.1.e1]
6. Changizi M, Kaveh MH. Effectiveness of the mHealth technology in improvement of healthy behaviors in an elderly population—a systematic review. mHealth 2017 Nov 27;3:51-51. [doi: 10.21037/mhealth.2017.08.06]
7. Whitehead L, Seaton P. The Effectiveness of Self-Management Mobile Phone and Tablet Apps in Long-term Condition Management: A Systematic Review. J Med Internet Res 2016 May 16;18(5):e97. [doi: 10.2196/jmir.4883]
8. Oliveira JGRD, Askari M, Silva Junior GBD, Freitas Filho RAD, Vasconcelos Filho JE. Renal Health: An Innovative Application to Increase Adherence to Treatment Through Self-monitoring for Patients With CKD and Provide Information for the General Population. Kidney International Reports 2019 Apr;4(4):609-613. [doi: 10.1016/j.ekir.2019.01.008]
9. Helbostad J, Vereijken B, Becker C, Todd C, Taraldsen K, Pijnappels M, et al. Mobile Health Applications to Promote Active and Healthy Ageing. Sensors 2017 Mar 18;17(3):622. [doi: 10.3390/s17030622]
10. Chiarini G, Ray P, Akter S, Masella C, Ganz A. mHealth Technologies for Chronic Diseases and Elders: A Systematic Review. IEEE J. Select. Areas Commun 2013 Sep;31(9):6-18. [doi: 10.1109/jsac.2013.sup.0513001]
11. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile Health Applications for the Most Prevalent Conditions by the World Health Organization: Review and Analysis. J Med Internet Res 2013 Jun 14;15(6):e120. [doi: 10.2196/jmir.2600]
12. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M, Sainz-De-Abajo B. Comparison of Mobile Apps for the Leading Causes of Death Among Different Income Zones: A Review of the Literature and App Stores. JMIR Mhealth Uhealth 2014 Jan 09;2(1):e1. [doi: 10.2196/mhealth.2779]
13. Akiyama M, Abraham C. Comparative cost-benefit analysis of tele-homecare for community-dwelling elderly in Japan: Non-Government versus Government Supported Funding Models. International Journal of Medical Informatics 2017 Aug;104:1-9. [doi: 10.1016/j.ijmedinf.2017.04.017]
14. Drummond M, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes. USA: Oxford University Press; 2015.
15. Bergmo TS. How to Measure Costs and Benefits of eHealth Interventions: An Overview of Methods and Frameworks. J Med Internet Res 2015 Nov 09;17(11):e254. [doi: 10.2196/jmir.4521]
16. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):1-7. [doi: 10.1186/s12874-018-0611-x]
17. Bramer WM, Giustini D, Kramer BM, Anderson P. The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. Syst Rev 2013 Dec 23;2(1):1-9. [doi: 10.1186/2046-4053-2-115]
18. Bramer WM, De Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. jmla 2018 Oct 04;106(4):531. [doi: 10.5195/jmla.2018.283]
19. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012:276-282. [doi: 10.11613/bm.2012.031]

20. Dimmick SL, Mustaleski C, Burgiss SG, Welsh T. Home Healthcare Nurse 2000;18(2):124-135. [doi: 10.1097/00004045-200002000-00013]
21. Lustbader D, Mudra M, Romano C, Lukoski E, Chang A, Mittelberger J, et al. The Impact of a Home-Based Palliative Care Program in an Accountable Care Organization. Journal of Palliative Medicine 2017 Jan;20(1):23-28. [doi: 10.1089/jpm.2016.0265]
22. Walsh M, Coleman JR. Developing a Pilot Telehealth Program. Home Healthcare Nurse: The Journal for the Home Care and Hospice Professional 2005;23(3):188-191. [doi: 10.1097/00004045-200503000-00014]
23. Walsh M, Coleman JR. Trials and Tribulations: A Small Pilot Telehealth Home Care Program for Medicare Patients. Geriatric Nursing 2005 Nov;26(6):343-346. [doi: 10.1016/j.gerinurse.2005.09.021]
24. Henderson C, Knapp M, Fernández JL, Beecham J, Hirani SP, Cartwright M, Whole System Demonstrator evaluation team. Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial. BMJ 2013 Mar 28;346(mar28 6):f2065-f2065. [doi: 10.1136/bmj.f2065]
25. Hicks LL, Fleming DA, Desaulnier A. The Application of Remote Monitoring to Improve Health Outcomes to a Rural Area. Telemedicine and e-Health 2009 Sep;15(7):664-671. [doi: 10.1089/tmj.2009.0009]
26. Rantz M, Lane K, Phillips LJ, Despins LA, Galambos C, Alexander GL, et al. Enhanced registered nurse care coordination with sensor technology: Impact on length of stay and cost in aging in place housing. Nursing Outlook 2015 Nov;63(6):650-655. [doi: 10.1016/j.outlook.2015.08.004]
27. Bendixen RM, Levy CE, Olive ES, Kobb RF, Mann WC. Cost Effectiveness of a Telerehabilitation Program to Support Chronically Ill and Disabled Elders in Their Homes. Telemedicine and e-Health 2009 Jan;15(1):31-38. [doi: 10.1089/tmj.2008.0046]
28. Noel HC, Vogel DC, Erdos JJ, Cornwall D, Levin F. Home Telehealth Reduces Healthcare Costs. Telemedicine Journal and e-Health 2004 Jun;10(2):170-183. [doi: 10.1089/tmj.2004.10.170]
29. Paré G, Poba-Nzaou P, Sicotte C. Home telemonitoring for chronic disease management: an economic assessment. Int J Technol Assess Health Care 2013 Mar 20;29(2):155-161. [doi: 10.1017/s0266462313000111]
30. Finkelstein SM, Speedie SM, Potthoff S. Home Telehealth Improves Clinical Outcomes at Lower Cost for Home Healthcare. Telemedicine and e-Health 2006 Apr;12(2):128-136. [doi: 10.1089/tmj.2006.12.128]
31. Johnston B, Weeler L, Deuser J, Sousa KH. Outcomes of the kaiser permanente tele-home health research project. Arch Fam Med 2000;9(1):40-45. [doi: 10.1001/archfami.9.1.40]
32. Folino AF, Breda R, Calzavara P, Borghetti F, Comisso J, Iliceto S, et al. Remote follow-up of pacemakers in a selected population of debilitated elderly patients. Europace 2012 Nov 01;15(3):382-387. [doi: 10.1093/europace/eus351]
33. Blum K, Gottlieb SS. The Effect of a Randomized Trial of Home Telemonitoring on Medical Costs, 30-Day Readmissions, Mortality, and Health-Related Quality of Life in a Cohort of Community-Dwelling Heart Failure Patients. Journal of Cardiac Failure 2014 Jul;20(7):513-521. [doi: 10.1016/j.cardfail.2014.04.016]
34. Liu SX, Xiang R, Lagor C, Liu N, Sullivan K. Economic Modeling of Heart Failure Telehealth Programs: When Do They Become Cost Saving? International Journal of Telemedicine and Applications 2016;2016:1-9. [doi: 10.1155/2016/3289628]
35. Grustam AS, Severens JL, De Massari D, Buyukkaramikli N, Koymans R, Vrijhoef HJ. Cost-Effectiveness Analysis in Telehealth: A Comparison between Home Telemonitoring, Nurse Telephone Support, and Usual Care in Chronic Heart Failure Management. Value in Health 2018 Jul;21(7):772-782. [doi: 10.1016/j.jval.2017.11.011]
36. Hwang R, Morris NR, Mandrusiak A, Bruning J, Peters R, Korczyk D, et al. Cost-Utility Analysis of Home-Based Telerehabilitation Compared With Centre-Based Rehabilitation in Patients With Heart Failure. Heart, Lung and Circulation 2019 Dec;28(12):1795-1803. [doi: 10.1016/j.hlc.2018.11.010]
37. Jerant AF, Azari R, Nesbitt TS. Reducing the Cost of Frequent Hospital Admissions for Congestive Heart Failure. Medical Care 2001;39(11):1234-1245. [doi: 10.1097/00005650-200111000-00010]
38. Myers S, Grant RW, Lugn NE, Holbert B, Kvedar JC. Impact of Home-Based Monitoring on the Care of Patients with Congestive Heart Failure. Home Health Care Management & Practice 2016 Jun 30;18(6):444-451. [doi: 10.1177/1084822306289991]
39. Boyne JJ, Di Van Asselt A, Gorgels AP, Steuten LM, De Weerd G, Kragten J, et al. Cost–effectiveness analysis of telemonitoring versus usual care in patients with heart failure: The TEHAF–study. J Telemed Telecare 2016 Jul 09;19(5):242-248. [doi: 10.1177/1357633x13495478]
40. Cui Y, Doupe M, Katz A, Nyhof P, Forget E. Economic Evaluation of Manitoba Health Lines in the Management of Congestive Heart Failure. hcpol 2013 Nov 27;9(2):36-50. [doi: 10.12927/hcpol.2013.23620]
41. Thokala P, Baalbaki H, Brennan A. Telemonitoring After Discharge from Hospital with Heart Failure – Cost-Effectiveness Modelling of Alternative Service Designs. Value in Health 2013 Nov;16(7):A530. [doi: 10.1016/j.jval.2013.08.1305]
42. Nouryan CN, Morahan S, Pecinka K, Akerman M, Lesser M, Chaikin D, et al. Home Telemonitoring of Community-Dwelling Heart Failure Patients After Home Care Discharge. Telemedicine and e-Health 2019 Jun;25(6):447-454. [doi: 10.1089/tmj.2018.0099]

43. Burri H, Sticherling C, Wright D, Makino K, Smala A, Tilden D. Cost-consequence analysis of daily continuous remote monitoring of implantable cardiac defibrillator and resynchronization devices in the UK. Europace 2013 Apr 18;15(11):1601-1608. [doi: 10.1093/europace/eut070]
44. Dixon P, Hollinghurst S, Edwards L, Thomas C, Gaunt D, Foster A, et al. Cost-effectiveness of telehealth for patients with raised cardiovascular disease risk: evidence from the Healthlines randomised controlled trial. BMJ Open 2016 Aug 26;6(8):e012352. [doi: 10.1136/bmjopen-2016-012352]
45. Capucci A, De Simone A, Luzi M, Calvi V, Stabile G, D'Onofrio A, et al. Economic impact of remote monitoring after implantable defibrillators implantation in heart failure patients: An analysis from the EFFECT study. Europace 2017;19(9):1493-1499. [doi: 10.1093/europace/eux017]
46. Perl S, Stiegler P, Rotman B, Prenner G, Lercher P, Anelli-Monti M, et al. Socio-economic effects and cost saving potential of remote patient monitoring (SAVE-HM trial). International Journal of Cardiology 2013 Nov;169(6):402-407. [doi: 10.1016/j.ijcard.2013.10.019]
47. Parahuleva MS, Soydan N, Divchev D, Lüsebrink U, Schieffer B, Erdogan A. Home monitoring after ambulatory implanted primary cardiac implantable electronic devices: The home ambulance pilot study. Clin Cardiol 2017 Aug 21;40(11):1068-1075. [doi: 10.1002/clc.22772]
48. Chen Y, Lin Y, Hung C, Huang C, Yeih D, Chuang P, et al. Clinical Outcome and Cost-Effectiveness of a Synchronous Telehealth Service for Seniors and Nonseniors with Cardiovascular Diseases: Quasi-Experimental Study. J Med Internet Res 2013 Apr 24;15(4):e87. [doi: 10.2196/jmir.2091]
49. Paré G, Poba-Nzaou P, Sicotte C, Beaupré A, Lefrançois É, Nault D, et al. Comparing the costs of home telemonitoring and usual care of chronic obstructive pulmonary disease patients: A randomized controlled trial. European Research in Telemedicine / La Recherche Européenne en Télémédecine 2013 Jun;2(2):35-47. [doi: 10.1016/j.eurtel.2013.05.001]
50. Witt Udsen F, Lilholt PH, Hejlesen O, Ehlers L. Cost-effectiveness of telehealthcare to patients with chronic obstructive pulmonary disease: results from the Danish 'TeleCare North' cluster-randomised trial. BMJ Open 2017 May 17;7(5):e014616. [doi: 10.1136/bmjopen-2016-014616]
51. Paré G, Sicotte C, St.-Jules D, Gauthier R. Cost-Minimization Analysis of a Telehomecare Program for Patients with Chronic Obstructive Pulmonary Disease. Telemedicine and e-Health 2006 Apr;12(2):114-121. [doi: 10.1089/tmj.2006.12.114]
52. Haesum LK, Soerensen N, Dinesen B, Nielsen C, Grann O, Hejlesen O, et al. Cost-Utility Analysis of a Telerehabilitation Program: A Case Study of COPD Patients. Telemedicine and e-Health 2012 Nov;18(9):688-692. [doi: 10.1089/tmj.2011.0250]
53. Clarke M, Fursse J, Brown-Connolly NE, Sharma U, Jones R. Evaluation of the National Health Service (NHS) Direct Pilot Telehealth Program: Cost-Effectiveness Analysis. Telemedicine and e-Health 2018 Jan;24(1):67-76. [doi: 10.1089/tmj.2016.0280]
54. Achelrod D, Schreyögg J, Stargardt T. Health-economic evaluation of home telemonitoring for COPD in Germany: evidence from a large population-based cohort. Eur J Health Econ 2016 Oct 3;18(7):869-882. [doi: 10.1007/s10198-016-0834-x]
55. Vitacca M, Bianchi L, Guerra A, Fracchia C, Spanevello A, Balbi B, et al. Tele-assistance in chronic respiratory failure patients: a randomised clinical trial. European Respiratory Journal 2008 Sep 17;33(2):411-418. [doi: 10.1183/09031936.00005608]
56. De San Miguel K, Smith J, Lewin G. Telehealth Remote Monitoring for Community-Dwelling Older Adults with Chronic Obstructive Pulmonary Disease. Telemedicine and e-Health 2013 Sep;19(9):652-657. [doi: 10.1089/tmj.2012.0244]
57. Mirón Rubio M, Ceballos Fernández R, Parras Pastor I, Palomo Iloro A, Fernández Félix BM, Medina Miralles J, et al. Telemonitoring and home hospitalization in patients with chronic obstructive pulmonary disease: study TELEPOC. Expert Review of Respiratory Medicine 2018 Feb 21;12(4):335-343. [doi: 10.1080/17476348.2018.1442214]
58. Dinesen B, Haesum LK, Soerensen N, Nielsen C, Grann O, Hejlesen O, et al. Using preventive home monitoring to reduce hospital admission rates and reduce costs: a case study of telehealth among chronic obstructive pulmonary disease patients. J Telemed Telecare 2012 May 31;18(4):221-225. [doi: 10.1258/jtt.2012.110704]
59. Le Goff-Pronost M, Mourgeon B, Blanchère J, Teot L, Benateau H, Dompmartin A. Real-world clinical evaluation and costs of telemedicine for chronic wound management. Int J Technol Assess Health Care 2018 Oct 29;34(6):567-575. [doi: 10.1017/s0266462318000685]
60. Binder B, Hofmann-Wellenhof R, Salmhofer W, Okcu A, Kerl H, Soyer HP. Teledermatological Monitoring of Leg Ulcers in Cooperation With Home Care Nurses. Arch Dermatol 2007 Dec 01;143(12):1511-1514. [doi: 10.1001/archderm.143.12.1511]
61. Dansky KH, Palmer L, Shea D, Bowles KH. Cost Analysis of Telehomecare. Telemedicine Journal and e-Health 2001 Sep;7(3):225-232. [doi: 10.1089/153056201316970920]
62. Wittenborn JS, Clemons T, Regillo C, Rayess N, Liffmann Kruger D, Rein D. Economic Evaluation of a Home-Based Age-Related Macular Degeneration Monitoring System. JAMA Ophthalmol 2017 May 01;135(5):452. [doi: 10.1001/jamaophthalmol.2017.0255]
63. Tousignant M, Moffet H, Nadeau S, Mérette C, Boissy P, Corriveau H, et al. Cost Analysis of In-Home Telerehabilitation for Post-Knee Arthroplasty. J Med Internet Res 2015 Mar 31;17(3):e83. [doi: 10.2196/jmir.3844]
64. Dorsey ER, Venkataraman V, Grana MJ, Bull MT, George BP, Boyd CM, et al. Randomized controlled clinical trial of. JAMA Neurol 2013 May;70(5):565-570 [FREE Full text] [doi: 10.1001/jamaneurol.2013.123] [Medline: 23479138]

65. Aoki N, Ohta S, Yamamoto H, Kikuchi N, Dunn K. Triangulation Analysis of Tele-Palliative Care Implementation in a Rural Community Area in Japan. Telemedicine and e-Health 2006 Dec;12(6):655-662. [doi: 10.1089/tmj.2006.12.655]
66. Vincent C, Reinharz D, Deaudelin I, Garceau M, Talbot LR. Public telesurveillance service for frail elderly living at home, outcomes and cost evolution: a quasi experimental design with two follow-ups. Health Qual Life Outcomes 2006 Jul 07;4(1):1-10. [doi: 10.1186/1477-7525-4-41]
67. Finch M, Griffin K, Pacala JT. Reduced Healthcare Use and Apparent Savings with Passive Home Monitoring Technology: A Pilot Study. J Am Geriatr Soc 2017 Apr 13;65(6):1301-1305. [doi: 10.1111/jgs.14892]
68. Upatising B, Wood DL, Kremers WK, Christ SL, Yih Y, Hanson GJ, et al. Cost Comparison Between Home Telemonitoring and Usual Care of Older Adults: A Randomized Trial (Tele-ERA). Telemedicine and e-Health 2015 Jan;21(1):3-8. [doi: 10.1089/tmj.2014.0021]
69. Magnusson L, Hanson E. Supporting frail older people and their family carers at home using information and communication technology: cost analysis. J Adv Nurs 2005 Sep;51(6):645-657. [doi: 10.1111/j.1365-2648.2005.03541.x]
70. Aanesen M, Lotherington AT, Olsen F. Smarter elder care? A cost-effectiveness analysis of implementing technology in elder care. Health Informatics J 2011 Sep 20;17(3):161-172. [doi: 10.1177/1460458211409716]
71. Dear BF, Zou JB, Ali S, Lorian CN, Johnston L, Sheehan J, et al. Clinical and Cost-Effectiveness of Therapist-Guided Internet-Delivered Cognitive Behavior Therapy for Older Adults With Symptoms of Anxiety: A Randomized Controlled Trial. Behavior Therapy 2015 Mar;46(2):206-217. [doi: 10.1016/j.beth.2014.09.007]
72. Nijhof N, Van Gemert-Pijnen LJ, Woolrych R, Sixsmith A. An evaluation of preventive sensor technology for dementia care. J Telemed Telecare 2013 Feb 22;19(2):95-100. [doi: 10.1258/jtt.2012.120605]
73. Titov N, Dear BF, Ali S, Zou JB, Lorian CN, Johnston L, et al. Clinical and Cost-Effectiveness of Therapist-Guided Internet-Delivered Cognitive Behavior Therapy for Older Adults With Symptoms of Depression: A Randomized Controlled Trial. Behavior Therapy 2015 Mar;46(2):193-205. [doi: 10.1016/j.beth.2014.09.008]
74. Bassi J, Lau F. Measuring value for money: a scoping review on economic evaluation of health information systems. J Am Med Inform Assoc 2013 Jul 01;20(4):792-801. [doi: 10.1136/amiajnl-2012-001422]
75. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS ONE 2017 Feb 2;12(2):e0170581. [doi: 10.1371/journal.pone.0170581]
76. Perry-Duxbury M, van Exel J, Brouwer W. How to value safety in economic evaluations in health care? A review of applications in different sectors. Eur J Health Econ 2019 Jun 6;20(7):1041-1061. [doi: 10.1007/s10198-019-01076-9]
77. Askari M, Klaver NS, van Gestel TJ, van de Klundert J. Intention to use Medical Apps Among Older Adults in the Netherlands: Cross-Sectional Study. J Med Internet Res 2020 Sep 04;22(9):e18080 [FREE Full text] [doi: 10.2196/18080] [Medline: 32624465]

Abbreviations

CBA: cost-benefit analysis
CEA: cost-effectiveness analysis
CUA: cost utility analysis
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analysis
R&D: research and development

Edited by G Eysenbach; submitted 18.09.20; peer-reviewed by T Kindratt, P Winkler, V Minichiello; comments to author 30.10.20; revised version received 04.12.20; accepted 13.01.21; published 09.03.21.

Please cite as:
Sülz S, van Elten HJ, Askari M, Weggelaar-Jansen AM, Huijsman R
eHealth Applications to Support Independent Living of Older Persons: Scoping Review of Costs and Benefits Identified in Economic Evaluations
J Med Internet Res 2021;23(3):e24363
URL: https://www.jmir.org/2021/3/e24363
doi:10.2196/24363
PMID:33687335

Viewpoint

Effort-Optimized Intervention Model: Framework for Building and Analyzing Digital Interventions That Require Minimal Effort for Health-Related Gains

Amit Baumel[1*], PhD; Frederick J Muench[2*], PhD

[1]Department of Community Mental Health, University of Haifa, Haifa, Israel

[2]Partnership to End Addiction, New York, NY, United States

[*]all authors contributed equally

Corresponding Author:
Frederick J Muench, PhD
Partnership to End Addiction
711 Third Avenue
New York, NY, 10017
United States
Phone: 1 9175320623
Email: fmuench@toendaddiction.org

Abstract

The majority of digital health interventions lean on the promise of bringing health and self-care into people's homes and hands. However, these interventions are delivered while people are in their triggering environments, which places competing demands on their attention. Individuals struggling to change or learn a new behavior have to work hard to achieve even a minor change because of the automatic forces propelling them back to their habitual behaviors. We posit that effort and burden should be explored at the outset and throughout the digital intervention development process as a core therapeutic mechanism, beyond the context of design or user experience testing. In effort-focused conceptualization, it is assumed that, even though goals are rational and people want to achieve them, they are overtaken by competing cognitive, emotional, and environmental processes. We offer the term effort-optimized intervention to describe interventions that focus on user engagement in the face of competing demands. We describe design components based on a 3-step process for planning an effort-optimized intervention: (1) nurturing effortless cognitive and environmental salience to help people keep effort-related goals prominent despite competition; (2) making it as effortless as possible to complete therapeutic activities to avoid ego depletion and self-efficacy reduction; and (3) turning the necessary effortful activities into sustainable assets. We conclude by presenting an example of designing a digital health intervention based on the effort-optimized intervention model.

(J Med Internet Res 2021;23(3):e24905) doi:10.2196/24905

KEYWORDS

behavior change; digital health; mental health; addiction; intervention; behavioral health; effort; salience; persuasive design

Introduction

Effort is physical, mental, or emotional exertion in an attempt to meet a goal. The effort exerted by an individual depends on the interplay between internal (eg, cognitive ability, motivation) and external (eg, social and environmental) facilitators and the barriers between an individual and their desired objective [1,2]. Substantial research has highlighted that reducing the effort needed to achieve an objective will increase the likelihood of achieving that objective in contexts ranging from consumer behavior to friendship [3]. The modern consumer technology industry has essentially been built on the premise of reducing the complexity and number of steps needed to reach desired objectives. Subsequently, the majority of digital health interventions have leaned on the promise of bringing health and self-care into people's homes and hands, overcoming the barriers to traditional services such as distance to a clinic, transportation, childcare, and more [4-8].

Despite the fact that digital interventions have significantly expanded their reach, with tens of millions of app downloads, the retention rates across the range of digital interventions remain very poor, with only 4% of behavioral health app users continuing use after 15 days [9]. Subsequently, findings suggest that user engagement with a digital intervention is 4 times higher

under trial in comparison with the use of the same intervention in the real world [10]. A recent report [11] on the use of MindSpot, an Australian digital mental health service may shed light on this phenomenon—the researchers reported an increase in the proportion of users looking for confidential assessment and a substantial decrease in the proportion of users looking for a traditional course-based internet intervention [11]. This suggests that many users are expecting far shorter therapeutic encounters such as microinterventions [12] compared with what would have been traditionally expected from users engaging digital health interventions.

We propose that a primary challenge with user engagement in digital interventions is that individuals who are struggling to change must work hard to achieve even a minor change because of the automatic forces propelling them back to their habitual behaviors [13-15]. Substantial literature has emphasized the continuum of automatic processes driving psychological distress and effortful processing fostering psychological health [16,17]. Not surprisingly, those with severe addiction and mental health disorders typically require a higher level of care (eg, inpatient care) to reduce the severity of symptoms in a controlled environment where recovery is the most salient cue. In effect, individuals have the headspace to work on their goals without being bombarded by environmental cues [18].

An effort-focused intervention model changes our conceptualization in the sense that we assume that goals are rational and that people want to achieve them. However, competing events in people's lives either require less effort or are more salient. As the next generation of digital health interventions is developed, we argue that an exploration of effort and burden should form the baseline for intervention development.

Existing Literature on Effort Reduction

The fields of user experience, heuristic evaluation, and persuasive design focus on principles such as user control, simplicity, predictability, and satisfaction specifically designed to increase engagement [19-21]. The outputs of these efforts range from autofill opportunities to 1-click shopping and frictionless feeds. In behavioral economics, effort reduction is often achieved with a default option [22]. For example, in their seminal paper on organ donation, Johnson and Goldstein [23] posited that one of the mechanisms of increased donations is that "making a decision often involves effort, whereas accepting the default is effortless." Environmental engineering theory, popularized by books [24], and seminal studies [25] on manipulating availability and access to different foods have revealed that reducing cognitive effort by making healthier choices available and unhealthy choices more burdensome to obtain improves healthy behaviors significantly and unconsciously. Underlying gamification presents perceived effort reduction by enhancing reward and reinforcement while pursuing a goal in a fun and engaging way [26]. Subsequently, a recent review [27] has shown that clinical applications that reduce the effort required from participants to engage in a desired response decrease self-injurious behavior, decrease pica, and increase appropriate eating.

These approaches are used often in the digital behavior change, supplemented by targeted persuasive intervention design for behavior change. For example, the Fogg model introduced the concept of the trigger into social-cognitive theory [28,29], that is, triggers presented at the right time in the right context reduce cognitive effort and increase motivation. More comprehensive taxonomies and persuasive models have been developed to identify core elements of behavior change interventions that drive engagement. For example, Oinas-Kukkonen and Harjumaa [30] developed a set of principles to build sustainable interventions that include concepts such as tunneling and choice reduction to foster engagement. Michie and colleagues [31] have developed a set of core behavior change principles, in which effort-reduction is implied, to guide intervention development heavily focused on learning theory and shaping behavior.

One of the reasons text-messaging interventions are acceptable may not be because they are just-in-time interventions, but rather, because individuals do not have to do anything except passively receive a text message once they sign-up. Text-messaging interventions have higher engagement over time than app-based interventions for perhaps no other reason than their effortlessness. For example, after 10 months of being signed up for the Text4Baby SMS intervention, 74.4% of mothers were still receiving messages [32]. To further increase sustained engagement the study [32] reported that "the extra step required to update the service with the birth date is being removed in case this has been a barrier to maintaining participation."

We are not positing that the focus on effort reduction is a new phenomenon. We are suggesting that effort reduction is often overlooked by our field as we develop interventions from the outset and at every stage of intervention engagement. The theories described above such as tunneling, persuasive design, and gamification are methods that reduce effort as a passive result of the optimal state rather than by the design goal of fostering such a state. If we focus on gamification alone, for example, we may miss opportunities for effort reduction at every stage of the behavior change process; however, if we focus on effort reduction, gamification will likely be included as one task within a larger effort-optimized intervention model.

Effort Optimized Intervention Model: Fostering Effortful Behavior by Making it as Effortless as Possible

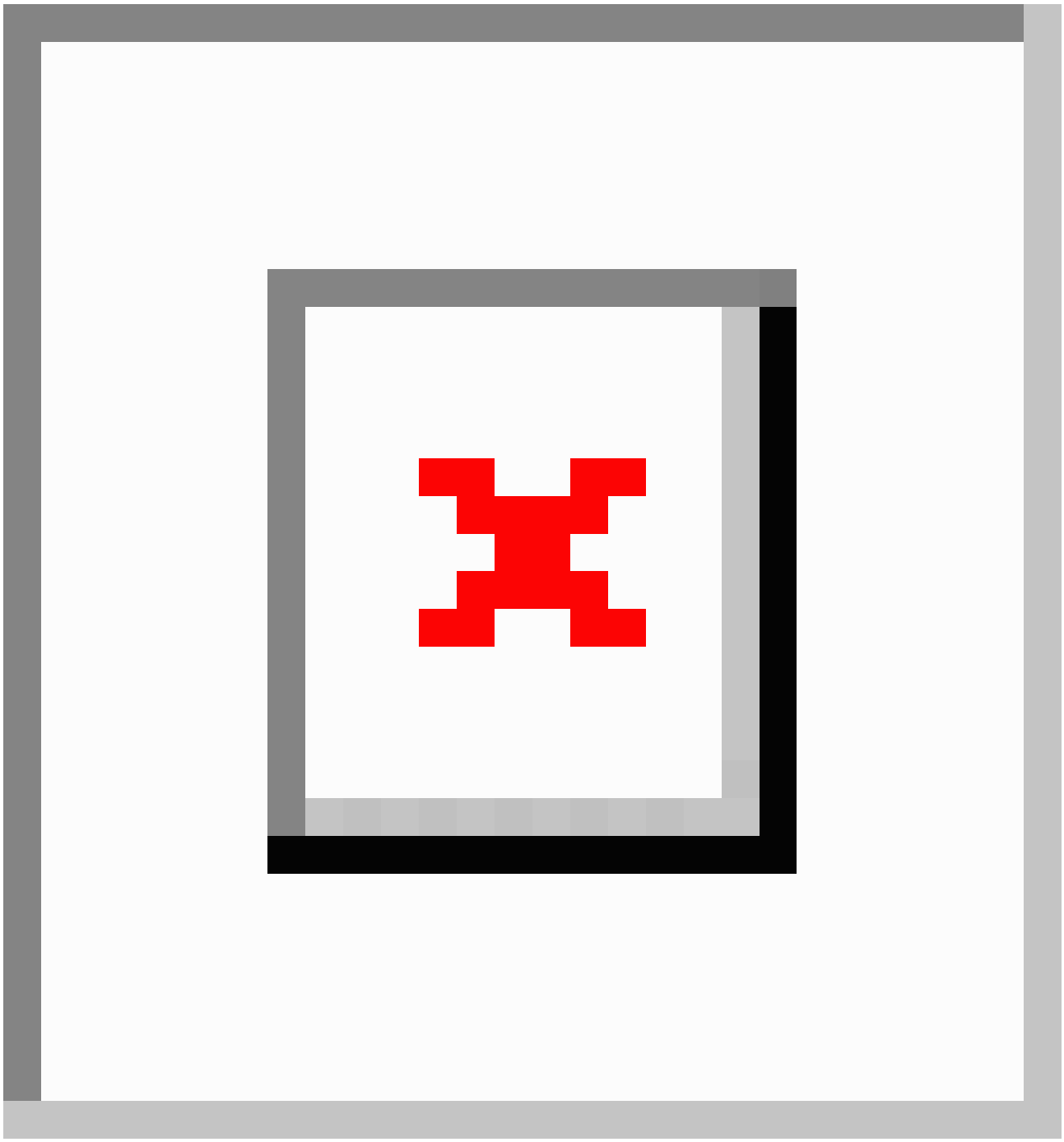
We offer the term *effort-optimized intervention* to describe interventions that focus on generating engagement with processes of therapeutic change in the face of competing demands. Here *engagement* refers to the time window of the intervention itself which may vary—mostly stretching from days to months [12]—and refers to the notion that the user has to engage with the intervention for the targeted time window for it to reach a desired impact. Understanding effort optimization starts with the question "what is the lowest burden method to trigger behavior change?" For example, if one is

trying to reduce arousal before bedtime, interventions may require a range of engagement levels (Figure 1).

Exploring the continuum of required effort enables the selection of interventions that meet individuals where they are in terms of motivation, ability, and barriers [28]. A person with almost no self-efficacy in changing a behavior may easily change the display options on their mobile phone to reduce blue light after 8 PM but may be unlikely to engage in guided paced breathing. At the same time, there may be no barriers to creating additive models of effortful engagement for those who are motivated and engaged. Unfortunately, the majority of effort targets for behavioral and mental health have fallen on the higher end of the spectrum. As a result, we are required to optimize effortful behaviors in times when we cannot make them fully effortless or passive.

We describe a 3-step process in the design of an effort-optimized intervention sequence, involving (1) nurturing salience to increase the chance of desired behaviors occurring in the face of competition; (2) making it as effortless as possible to complete therapeutic activities in order to avoid ego depletion and self-efficacy reductions; and (3) turning the necessary effortful activities into sustainable assets.

Figure 1. Effortless to effortful intervention examples.



Nurturing Salience to Increase the Chance of Desired Behaviors Occurring in the Face of Competition

Background: What Makes a Therapeutic Target More Salient?

Salience refers to how much a certain object, either internal or external, is prominent in one's mind. Salience can be triggered via a range of experiences—from an intense emotional event that becomes deeply encoded in memory to ongoing subtle cues embedded in one's daily routines over long periods of time. For the purposes of this paper, we discuss the latter trigger. (We use the term *trigger* to refer to the broad category of digital stimuli designed to prompt desired actions and reactions from users [33].) An object's salience can be defined based on its availability, whether it is actionable, and how much it is linked to a reward. (Table 1 contains a summary of components described with the body of this manuscript.)

Table 1. Summary of the effort-optimized intervention 3-step design process and related components.

Component	Explanation	Example
Nurturing salience		
Increasing task or goal availability	How easy is it to think about the goal compared to competing demands at a desired time point?	Just-in-time text messaging or push notifications about the targeted task; using implementation intentions to mark dinner as an environmental cue for a parent to conduct a family gratitude exercise
Creating an actionable script	Triggering a step-by-step script to convey exactly what to do to foster the automaticity of the script during the task	Embedding a simple step-by-step app guidance for parents learning what to do in the face of their child's panic attack
Incentive salience for a task or goal	A cognitive process that includes an automatic motivational component that links a person's desires to a rewarding stimulus to create a feedback loop toward behavior change	Offering a meaningful immediate reward through the app such as celebration of a successful running exercise
Optimizing novelty	Avoiding habituation by not presenting similar stimuli over and over again and varying the affective impact on the individual	Changing the delivery medium of inspirational motivational messages from text, video, and audio across a program
Making the completion of therapeutic activities as effortless as possible		
Setting graded tasks	Determining small and achievable goals, and moving forward in small steps	A mobile app beginning running distance at 0.5 km and gradually stepping the user up to 5 km
Setting dynamically tailored tasks	Adapting to the user's state based on passive data tasks they care about and past failures and successes	When it takes more time for the user to acquire a skill, they receive additional features from the program prior to moving forward.
Reducing the effort required to engage in therapeutic activities	Keeping all relevant tools available in-house; making it as easy as possible to perform the activity	Taking a photo of a meal through the app which analyzes it to document calorie intake; automatically triggering changes to screen color temperature based on time of day
Turning effort into assets		
Documenting and reflecting on past effort-related activities in a meaningful way	Turning effort into assets by documenting and reflecting on aspects users care about during the therapeutic process; once assets are made, users are inclined to keep investing so that their assets will not go to waste	Presenting effortful activities the user conducted (eg, user reports on socializing with a friend) and how these activities are helpful (eg increase life satisfaction other time)
Turning effort into a meaningful narrative	Helping people acknowledge the link between the effort they just exerted and their commitment to the therapeutic process	Upon reporting a positive interaction with their child, parents are asked to celebrate investing effort in becoming better parents
Reframing effort as positive	Embedding a narrative in which the reward is the respect for asserting effort beyond skills acquisition	Encouraging users who finished an online learning on coping with depression by stressing out how this activity shows their commitment to feeling better

Availability is the ease with which one is able to think about the target object at a given time point. A simple way to manipulate an object's availability is to trigger it using just-in-time mobile reminders or environmental triggers [33]. Critically, availability can be manipulated cognitively by priming people to think in a certain manner at a given time point, thus creating automaticity [34]. Presenting a certain object can prime a goal-directed behavior in the direction of the desired target, whether it be by using words associated with homophobia to increase implicit antigay bias [35], holding a warm object to increase altruism [36], or using the words "substance abuser" or "person with a substance-use disorder" to manipulate individuals' assumptions about whether someone should go to jail or to treatment [37]. These examples are congruent with the notion that creating cognitive prominence can trigger goal-directed behaviors without having to overtly instruct someone to be more mindful of a goal. Availability can also be triggered overtly through motivational reminders and environmental cues embedded in just-in-time digital interventions [38]. The end goal is to increase goal availability

during an effortful decision by reducing the amount of effort needed to retrieve the information.

An object is actionable when the person knows exactly what to do to achieve the desired outcome and how to do it at a given time point. The steps must become salient so that some action can be taken when goal-striving is triggered. Implementation intentions are priming methods that create if–then statements to trigger attention for a future desired outcome by making the association between a trigger and the resulting step-by-step behavior more immediate and less effortful [39,40]. Implementation intentions consist of a basic 2-step process to increase actionable behavior toward a goal, for example, (1) "when I sit down for dinner" (trigger); (2) "I will ask everyone to talk about one thing they are grateful for before I put the first bite in my mouth" (script). Other examples include online graphic illustrations and scenario-based scripts [41] to accompany text guidance. Using graphics targets different memory mechanisms and can help make a script more accessible from multiple pathways.

The rewarding aspect, defined as *incentive salience*, is a cognitive process that includes an automatic motivational component that links a person's desires or actions to a rewarding stimulus [42]. Incentive salience creates a feedback loop by which promise of the reward drives a person's attention. When it comes to behavior change in a person's natural environment, we assume that the reward for maladaptive behaviors such as parents yelling at their kid to "shut up" will be immediate (silence). Changing a person's behavior to adopt better practices requires significant effort and the promise of long-term rewards (eg, reducing behavior problems). Technological advances such as immersive virtual reality experiences, neurostimulation, and even actively targeting incentive salience by manipulating immediate rewards can create a reward-based feedback loop for behavior change. Immediate rewards can be produced by rewarding the attempt at the behavior and not the outcome (eg, making sure the parents understand that they are being evaluated based on their responses and not based on their child's behaviors and immediately celebrating their successes in improving their daily practices).

When we teach or ask the user to conduct a new internal (eg, cognitive reframing) or external (eg, exposure) therapeutic process or activity, the quality of our digital message delivery also affects how salient the targeted process will be in the user's mind. The more immersive, tangible, relatable, and personally tailored the message is, the more salient the targeted process will be. For example, when teaching a user to conduct an exposure paradigm, using automated scenario-based learning with video tutorials and relatable figures will be more immersive cognitively than explanations with texts; a text correspondence through an automated system that asks several questions and then provides personalized feedback and personalized motivational messaging that are meaningful in one's life will be emotionally more salient than general statements. Furthermore, because stimulus quality plays a significant role in drawing the user's attention, we must think about how to avoid habituation by not repeatedly presenting similar stimuli [43]. In effect, using novelty, such as changing the delivery medium, message type, and content, is key to maintaining user attention over time.

Finally, because the developers' goal is to design an effort-related intervention sequence that is sustainable in people's lives, they have to think about embedding these activities in an environmental context that will then serve as a natural environmental cue [44]. This enables the desired activity to be automatically triggered without having to draw attention to it—a requirement in the first steps of process acquisition. For example, when a desired positive interaction is to share a funny story with their child, parents could be prompted to do this during dinner. In this case, the developer views dinner time as an environmental cue.

From a public health perspective, when an individual is required to engage with a digital health intervention, it also means that unhealthy cognitions become more salient in the individual's mind—they are available, rewarding, and actionable—otherwise, this person would not need an intervention. Therefore, examining the interaction between salience, effortful behavior, and motivation can help us to understand the type of salience manipulation needed in a particular intervention sequence.

Promoting Desired Activities

Figure 2 presents a model that describes the probability of an activity occurring in the face of competing activities as a function of effort, motivation, and salience. This conceptualization follows Fogg's [28] work on determining the probability of a behavior occurring following a trigger based on the relationship between ability and motivation. We use effort instead of ability to stress the importance of subjective experience, which can fluctuate mainly due to levels of effort expenditure prior to a task and the available ego strength available to complete a task.

As shown in Figure 2, the probability of a behavior occurring is based on the relationship between effort and motivation. Activities located on the same curve have the same probability of occurring either because they are less effortful or because they are more motivating. Furthermore, if 2 prompted events compete over resources (eg, whether parents either yell at their kids or take deep breaths and try to calmly educate them), the activities located on a higher curve have a higher probability of occurring (that is, B will have a higher chance of occurring than A). Salience plays a crucial role in this process. Manipulating the salience of the desired activity (ie, making it more available, rewarding, and actionable) has the potential to increase the chances of the activity occurring in the face of competition by making the activity either more motivating or less effortful.

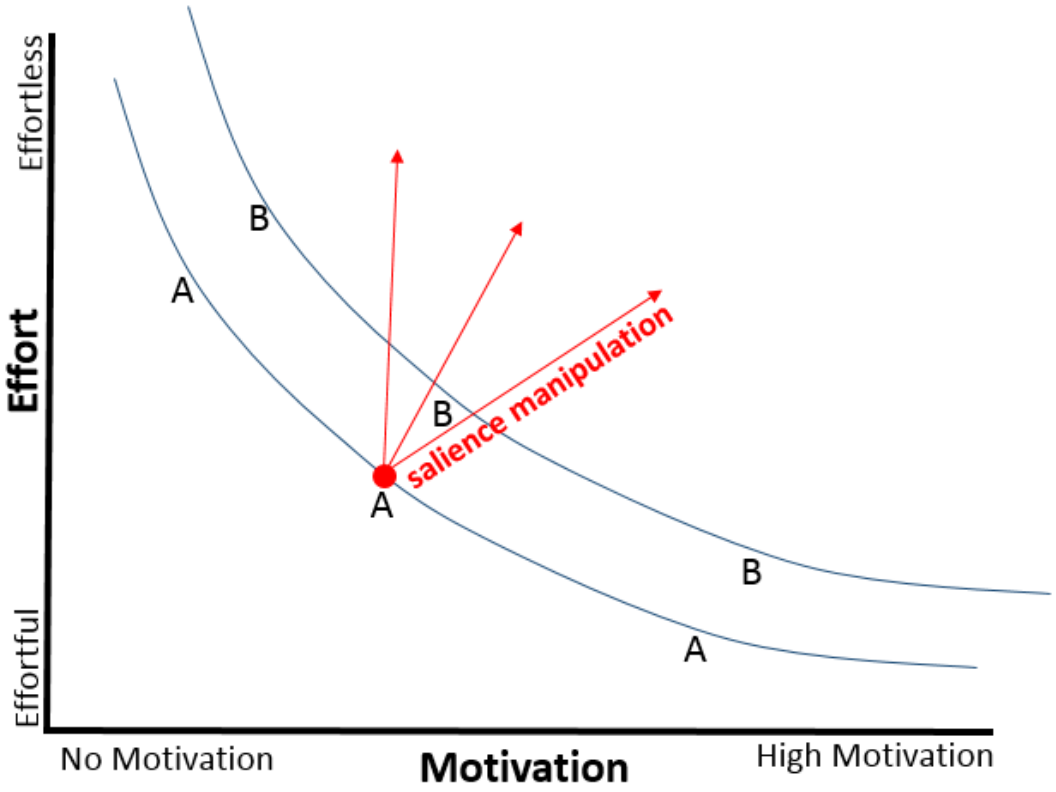
Availability increases the effortlessness of the targeted behavior because, in one's subjective experience, there are fewer competing or available activities. For example, having a playlist on the way home from work that includes a 1-minute audioclip that discusses the desired pre-evening activity makes it more available in the person's mind when they arrive home than other activities.

Actionability increases both effortlessness and motivation. For example, parents are presented with a tangible video that teaches them step-by-step what to do when their child misbehaves

(scenario-based learning), then they must confirm their understanding using a worksheet in which they write their own step-by-step process for the exact targeted behavior and print it out, and later that week, when their child acts in a certain way, they can easily identify the event and know exactly what to do. Consequently, they need to exert less effort to identify the trigger and decide on the action.

Incentive salience involves making the reward clear, tangible, and relatable. For example, a parent drives home from work and is prompted to listen to a 1-minute motivational audioclip on the significance of playing together with their child—an audioclip that also directs them to reflect on their time playing with their parent and how meaningful it was. The novelty and emotional activation of this exercise increases the prominence of the reward and the availability of the desired behavior (which, as suggested, also reduces the effort exerted when performing this activity).

Figure 2. The probability of an activity occurring in the face of competition as a function of effort, motivation, and salience. The blue curves are probability curves.



Making the Completion of Therapeutic Activities as Effortless as Possible

Overview

Developing models that are available, actionable, and rewarding require some effort at the outset of the behavior change process in order to reduce the effort exerted during daily goal-directed behaviors. This approach must also be accompanied by making the tasks themselves as effortless as possible.

Setting Graded and Dynamically Tailored Therapeutic Activities as Targets

The literature points to the importance of setting graded tasks, determining small and achievable goals, and moving forward in small steps as the user succeeds in prior steps [45]. Breaking a distal goal into achievable subgoals increases self-efficacy [46], which has been shown to be an important factor in determining whether a task will be initiated and successfully performed. From a behavior change standpoint, when using graded tasks, less effort is required for each activity. This increases the user's ability to engage in the task, thereby increasing the chances that the triggered activity will occur [28,40,44]. Critically, the use of graded tasks also involves improvement of the user's skills or condition in such a way that the next task becomes less effortful to achieve, as highlighted in shaping paradigms [47]. For example, when an untrained user is triggered by an app to run 5 km for the first time, the amount of effort required to complete this activity may be very high. However, if the user is triggered to run 1 km several times and then 3 km in a graded manner, the amount of effort related to running 5 km might be as little as it was to run 1 km for the first time.

One psychological technique that is not very common in digital interventions (though it could be easily incorporated) is the presentation of an artificial step prior to a subsequent step that otherwise might demand too much effort to complete. For example, when parents are taught to present an appropriate nonharsh consequence in the event of their child's disobedience [48,49], this step may feel like a giant leap, thus demanding plenty of effort in the parents' mind. We may, therefore, create another step of setting expectations for which parents are

directed to sit with their child prior to changing the way they react to them and to simply present the fact that they are going to do whatever they can to help them, meaning the rules of the house are going to change. This strategy enables parents to acquire self-efficacy in a small manner prior to anything else and could be easily incorporated in an adaptive digital program.

As implied, the adaptive nature of the tasks involves taking the user's state into account. User level of motivation is a moderating factor in determining the amount of effort the person is able to exert. Motivation is expected to fluctuate during the intervention based on prior successes and failures [50]. Users who encounter difficulties may require a different task or path than those who found the task easy to complete based on the interplay between self-efficacy, motivation, and goal achievement. Monitoring activities and user condition will enable the task to be dynamically tailored such that the effort required at any given moment is adequate. For example, adaptive goal interventions change the goal based on the user's successes or failures in achieving the goal, such as by increasing the number of weekly drinks allowed in a drinking moderation intervention when goals are not being met or, conversely, by changing to abstinence if moderation is being met with repeated failures. The real-time adaptation of the digital intervention is effort-optimized to meet the user's goals, motivation, and commitment [51].

Reducing the Effort Required to Engage in Therapeutic Activities

Developers can help to make therapeutic activities less effortful by reducing the cognitive and environmental effort required to successfully complete each activity [52-54]. Keeping all the relevant tools for completing a targeted activity within a digital component reduces the cognitive and environmental effort needed to search for those tools elsewhere [55]. It will be easier for users to follow a diet if the diet app makes all the information about the diet, the availability of support groups, and tools to document calorie intake available in-house [55].

Another aspect to consider is whether each tool makes it as easy as possible to conduct the desired activity [56,57]. For example, it would be easier for users to document their calorie intake if they could simply take a photograph of the meal and have the caloric results calculated automatically. Similarly, when training a behavior increases performance, a system can help to reduce the effort required for the training by providing the means to rehearse it [30]. Think, for example, of a person who is trying to overcome social phobia through graded tasks and who is now being asked to chat with a person they do not know in a nonjudgmental environment. In such an instance, providing an option of a click-button within the digital platform to connect them with a trained peer [58] would directly reduce the effort required.

Turning Effort Into Assets

An effort-optimized intervention does not mean that there is no effort on the part of the user. Indeed, effort contributes to sustainable change because the effort people are choosing or willing to make and the way they perceive it substantially impact the therapeutic process, intervention gains, and future effort capacity. Humans think in narratives with players, good and evil, conflicts, and dramatic changes in the plot [59]. Stories create structure because they have an inner rationale that corresponds with the past-present-future tenses, which enables people to predict the future based on the past. Therefore, people are built to create meaning based on their present experiences in a way that rationally fits with their past story and future direction, as captured in their identity, role in the world, and desires [59].

Effort is a very important ingredient in this process because the effort exerted for an activity will be used to create a meaning that mostly fits within the story we tell about ourselves. Experiments on cognitive dissonance theory and placebo effects strengthen this notion by showing that people who are asked to put more effort into an assignment later perceive it as being more meaningful to them [60]; meanwhile, people who pay more find a placebo to be more helpful [61]. Furthermore, several studies have suggested that people prefer to exert effort on a task when they are motivated to enhance their feelings of relatability, ownership, and control over the potential outcome. A notable example involves a US company that produced instant cakes ("just add water") in the 1940s targeting women maintaining their households. The product did not sell very well until the company removed some ingredients from the mix, such as eggs and milk, which required the baker to do more work during the baking process. It seemed that, with the increased effort, the baker felt greater ownership over the result and more deserving of the compliments for their work [62]. This phenomenon is described by Ariely as the *Ikea effect*—a cognitive bias that leads us to place higher value on things that we help to create [62].

To summarize, effort is a crucial ingredient in the way we create meaning because the amount of effort we invest in an activity impacts the extent to which we build meaningful stories around that activity. Specifically, the more effort we invest in something, the more meaningful we find it and the more committed we are to it. To take a literary example, when the Little Prince tries to explain to the fox what makes his rose different from the thousands of roses that appear to be identical, "It is the time you have wasted for your rose that makes your rose so important," he asserts [63].

Understanding this dynamic is key, as interventions that focus only on effort reduction and do not help people to feel that they have choice, acknowledge their work, and therefore, create meaning around the effort-based behaviors may fail in helping people to stay on the beneficial pathway when new challenges arise. In digital health interventions, people's efforts can be translated into assets by helping them to acknowledge the meaning of their work. In this way, we reframe effort as something positive by stressing that the effort exerted shows the user's commitment to and ownership of the therapeutic process. This shift toward a growth mindset and meaning-based acceptance can be embedded in all our work to increase effortlessness and decrease efficacy reductions based on an outcome mindset.

People's efforts can be turned into investment by documenting the aspects they care about during the therapeutic process [55,64]. Because users have already invested in the activity and created some assets, they are more inclined to move forward and keep investing in this path so that their assets will not go to waste. We stress that the desired documentation should be connected to aspects that are highly meaningful in people's lives, mostly within a social context, such as the time they got to spend with loved ones because they successfully executed an intervention's task.

These three components of nurturing salience, reducing effort to engage in therapeutic activities, and shifting the meaning of effortful behavior to become an asset can be embedded in both new and existing digital interventions. Whether it be by creating a simple visual diagram of the goal of a lesson at the beginning of a module, playing music in the background randomly to keep users engaged while completing a task, or including a narrative of an effortful journey during periods of declining motivation, we can reduce the effort needed for users to achieve positive outcomes. In turn, this will enable users to achieve their goals without taking away from the core therapeutic skill components of many interventions.

Designing a Digital Health Intervention Sequence Based on the Effort-Optimized Intervention Model

To further clarify the effort-optimized intervention model and how to design a digital intervention sequence accordingly, we provide an example using common intervention content in parent training programs for young children with disruptive behavior disorders [65-67] aimed at increasing the positive interactions between parent and child. For brevity, we only discuss the aspects of effort optimization during skill acquisition time, not other important aspects such as persuasive design or the therapeutic alliance nurtured between the user and the program [52,68-70]. Our baseline is a standard digital parent training intervention in which parents complete a short interactive module about positive parenting practices. Parents are instructed to increase the positive interactions at home and then directed to the next module a week or two later, depending on success.

Planning an effort-optimized intervention begins with defining the task, considering how it might be perceived by the participant, and identifying competing activities or challenges. In our example, the task is to increase positive interactions when the parent and child are in the same surrounding (eg, at home). For parents, it can be difficult to foster positive interactions because it is not always highly enjoyable at first, especially for those who have not naturally exercised such practices before; parents might not have clear ideas about such interactions and how easily they can be incorporated on a daily basis. Furthermore, such interactions are not necessarily linked to a tangible clear reward. We offer a few common competing events which may require less effort than nurturing a positive interaction with the child (although they are not based on empirical studies, it is important to use clear examples here for didactic reasoning)—by letting a child play a mobile app or watch television which requires much less effort, parents may find that playing with their own mobile device requires less effort and is rewarding in the sense of passing the time, and while parents may have some house chores or work they can do later, finishing them early offers a clear reward. Based on these challenges, developers can use the effort-optimized intervention framework to increase the chances of the desired activity being completed in the face of competition. As shown in Table 2, each concept informs the design of the intervention in a way that is directly related to increasing the chances that users will reach their objective despite competition. First, realizing that parents may want to have positive interactions with their child but lack good ideas, we offer these ideas in a concrete way (eg, ideas for what to discuss in the evening). Second, as we believe that parents may find letting their child watch television to be more rewarding than interacting, we have to address this competing activity both directly by helping parents emotionally connect to the difference between the two activities and indirectly by making the desired activity more salient in their mind. Third, we must acknowledge the parents' effort in order to create assets that help them feel good about the investment they have made. These considerations result in many new features that are not incorporated in a standard online module-based training environment.

Table 2. Features increasing parent–child positive interactions based on the effort-optimized intervention framework.

Component	Plan
Nurturing salience	
Increasing the task's availability	Triggers with relevant content (eg, ideas for what to discuss during dinner) sent in the hour before parents are home from work
	Priming parents to ask themselves about opportunities for positive interactions in the face of competing events (eg, thinking that their child would prefer to watch television instead and so not trying) through consistent but variable triggers, such as text questions, motivational scripts, and other minimal cues
Rewarding/incentive salience	Triggers connecting tangible rewards to the desired activities: "Think of your best memories with your parent. You putting some effort into playing with your child is something that will be far more memorable to you and him/her than times when you both watched separate screens."
	Directing parents to celebrate their positive interactions with their kids and to report on it using a mobile app
	Rewarding consistent attempts at behavior over outcomes through the platform (eg, the outcome is engaging in the behavior, not their child's behavior)
Creating an actionable script	A tailored list of positive interactions with brief step-by-step instructions based on an online questionnaire parents were asked to complete.
Optimizing novelty	Sending all triggers above using different delivery mediums (text, audio, and video), timing (time of day, day, special events), and personas (instructor, peers, celebrity testimonial)
Embedding tasks based on natural environmental cues	Directing parents to find one positive activity to conduct during dinner, such as a gratitude exercise that can be trigged through the mobile device in the right time
Making the completion of therapeutic activities as effortless as possible	
Setting graded tasks[a]	Asking parents to pick their preferred activities from an list of relevant activities, which automatically creates their own table that is then available on the website and as a printed version
Setting dynamically tailored tasks[a]	Creating a task list based on efficacy and effort. For example, if the parents report very low efficacy or past failures, a first step may be directing parents to sit with their child when the child is watching television and initiating a conversation
Turning effort into assets	
Documenting and reflecting on past effort-related activities in a meaningful way	Documenting reports in an accumulated manner on the home page of the app or website that offers rewards based on the level of engagement (eg, the amount of quality time reported so far). If not engaged, simple motivational statements replace effortful behavior rewards
Turning effort into a meaningful narrative; reframing effort as positive	Implementing automated feedback, which presents a narrative of them doing whatever they can to be good parents. For example: "the effort you invested today in trying to play with your kid shows how well you are committed to improve your relationship. You should be proud of yourself."

[a]Reducing the effort required to engage in therapeutic activities is embedded in this component as well.

Further Considerations, Future Directions, and Conclusions

Research and implementation of the effort-optimized intervention model demand that considerable attention be paid to some specific aspects. From a theoretical perspective, we need to learn more about what prevents people from performing desired behaviors at the individual level, even when they want to achieve them [2,71]. Studying such instances will enable developers to design user-centric products with relevant effort-optimized intervention sequences. Another line of research could focus on how people sustain beneficial behaviors over time, and more precisely, when and how competing events emerge and what people experience at these times. This knowledge will enable us to understand whether new triggers need to be incorporated into a future time window in order to avoid depletion. Finally, we need to learn how to develop effort-optimized intervention sequences so that they are not intrusive and thus eventually diminish people's desire and tendency to self-manage their situation.

The mechanistic study of effort reduction has been explored more in the consumer social media and commerce sectors in the form of A/B testing paradigms. In such paradigms, small changes to the user experience are repeatedly tested to optimize engagement, as small changes often lead to massive shifts in engagement (eg, "like" button, frictionless feed, page load time). While there are significant differences among these activities that require little effort with little meaningful long-term reward—and potentially significant negative consequences over time—their success highlights that, to create positive change, our attention as interventionists needs to shift to increase the 3% to 6% engagement rate in health applications. This is especially true given that research has revealed very few differences in outcomes between interventions with differing content or behavior change targets [72]. Effort-optimized intervention paradigms are designed to ensure that equal weight is placed on the content of our interventions and on how we engage and sustain individuals using common processes that adapt to meet individual needs, both in terms of what people need and how they consume and integrate it into their lives.

Conflicts of Interest

AB has received payment for consulting from Pro-Change Behavior Systems. FM has several patents on using vibration, light and electrical stimulation to induce changes in one's cardiovascular system.

References

1. Ainslie G. Willpower with and without effort. Behav Brain Sci 2020 Aug 26:1-81. [doi: 10.1017/s0140525x20000357]
2. Kotabe HP, Hofmann W. On integrating the components of self-control. Perspect Psychol Sci 2015 Sep 17;10(5):618-638. [doi: 10.1177/1745691615593382] [Medline: 26386000]
3. Dooley R. Friction: The Untapped Force That Can Be Your Most Powerful Advantage. New-York: McGraw Hill Professional; 2019.
4. Mohr D, Ho J, Duffecy J, Baron K, Lehman K, Jin L, et al. Perceived barriers to psychological treatments and their relationship to depression. J Clin Psychol 2010;66(4):394-409. [doi: 10.1037/t30550-000]
5. Mohr DC, Hart SL, Howard I, Julian L, Vella L, Catledge C, et al. Barriers to psychotherapy among depressed and nondepressed primary care patients. Ann Behav Med 2006 Dec;32(3):254-258. [doi: 10.1207/s15324796abm3203_12] [Medline: 17107299]
6. Neuroscience of psychoactive substance use and dependence. World Health Organization. 2004. URL: https://www.who.int/substance_abuse/publications/en/Neuroscience_E.pdf [accessed 2021-03-04]
7. Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, Lepine J, WHO World Mental Health Survey Consortium. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. JAMA 2004 Jun 02;291(21):2581-2590. [doi: 10.1001/jama.291.21.2581] [Medline: 15173149]
8. Kazdin AE. Addressing the treatment gap: a key challenge for extending evidence-based psychosocial interventions. Behav Res Ther 2017 Jan;88:7-18. [doi: 10.1016/j.brat.2016.06.004] [Medline: 28110678]
9. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
10. Baumel A, Edan S, Kane J. Is there a trial bias impacting user engagement with unguided e-mental health interventions? a systematic comparison of published reports and real-world usage of the same programs. Transl Behav Med 2019 Nov 25;9(6):1020-1033. [doi: 10.1093/tbm/ibz147] [Medline: 31689344]
11. Titov N, Dear BF, Nielssen O, Wootton B, Kayrouz R, Karin E, et al. User characteristics and outcomes from a national digital mental health service: an observational study of registrants of the Australian MindSpot Clinic. The Lancet Digital Health 2020 Nov;2(11):e582-e593. [doi: 10.1016/s2589-7500(20)30224-7]
12. Baumel A, Fleming T, Schueller SM. Digital micro interventions for behavioral and mental health gains: core components and conceptualization of digital micro intervention care. J Med Internet Res 2020 Oct 29;22(10):e20631 [FREE Full text] [doi: 10.2196/20631] [Medline: 33118946]
13. Hartlage S, Alloy LB, Vázquez C, Dykman B. Automatic and effortful processing in depression. Psychol Bull 1993 Mar;113(2):247-278. [doi: 10.1037/0033-2909.113.2.247] [Medline: 8451334]
14. Neal DT, Wood W, Labrecque JS, Lally P. How do habits guide behavior? perceived and actual triggers of habits in daily life. J Exp Soc Psychol 2012 Mar;48(2):492-498. [doi: 10.1016/j.jesp.2011.10.011]
15. Neal DT, Wood W, Wu M, Kurlander D. The pull of the past: when do habits persist despite conflict with motives? Pers Soc Psychol Bull 2011 Nov 22;37(11):1428-1437. [doi: 10.1177/0146167211419863] [Medline: 21859902]
16. Kahneman D. Attention and Effort. Englewood Cliffs, New Jersey: Prentice-Hall; 1973.
17. Bruya B, Tang Y. Is attention really effort? revisiting Daniel Kahneman's influential 1973 book. Front Psychol 2018;9(6):1133-1141 [FREE Full text] [doi: 10.3389/fpsyg.2018.01133] [Medline: 30237773]
18. Bordnick PS, Schmitz JM. Cocaine craving: an evaluation across treatment phases. J Subst Abus 1998 Jan;10(1):9-17. [doi: 10.1016/s0899-3289(99)80136-9]
19. Lemieux C, McDonald D. Frictionless: Why the Future of Everything Will be Fast, Fluid, and Made Just for You. New York: Harper Collins; 2020.
20. Nielsen J. Enhancing the explanatory power of usability heuristics. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1994 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 24-28; Boston, Massachusetts p. 152-158. [doi: 10.1145/191666.191729]
21. Krug S. Don't Make Me Think!: A Common Sense Approach to Web Usability. Berkeley, California: Pearson Education India; 2000.
22. Johnson E, Goldstein D. Decisions by Default. Princeton, NJ: Princeton University Press; 2013.
23. Johnson E, Goldstein D. Do defaults save lives? Science 2003 Nov 21;302(5649):1338-1339. [doi: 10.1126/science.1091721] [Medline: 14631022]
24. Thaler R, Sunstein C. Nudge: Improving Decisions About Health, Wealth, and Happiness. New York: Penguin Books; 2009.

25. Wansink B, Sobal J. Mindless eating. Environ Behav 2016 Jul 26;39(1):106-123. [doi: 10.1177/0013916506295573]
26. Wiemeyer J, Nacke L, Moser C, Floyd MF. Player experience. In: Dörner R, Göbel S, Effelsberg W, Wiemeyer J, editors. Serious Games: Foundations, Concepts and Practice. Cham: Springer; 2016:243-271.
27. Wilder DA, Ertel HM, Cymbal DJ. A review of recent research on the manipulation of response effort in applied behavior analysis. Behav Modif 2020 Feb 28:145445520908509. [doi: 10.1177/0145445520908509] [Medline: 32107925]
28. Fogg B. A behavior model for persuasive design. 2009 Presented at: 4th international Conference on Persuasive Technology; April 26-29; California, USA. [doi: 10.1145/1541948.1541999]
29. Fogg BJ. Persuasive Technology: Using Computers to Change What We Think and Do. Amsterdam: Morgan Kaufmann Publishers; 2003.
30. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. CAIS 2009;24:1. [doi: 10.17705/1cais.02428]
31. Michie S, West R, Campbell R, Brown J, Gainforth H. ABC of behaviour change theories: an essential resource for researchers. In: Policy Makers and Practitioners. Silverback IS: Silverback Publishing; 2014:402.
32. Whittaker R, Matoff-Stepp S, Meehan J, Kendrick J, Jordan E, Stange P, et al. Text4baby: development and implementation of a national text messaging health information service. Am J Public Health 2012 Dec;102(12):2207-2213. [doi: 10.2105/ajph.2012.300736]
33. Muench F, Baumel A. More than a text message: dismantling digital triggers to curate behavior change in patient centered health interventions. J Med Internet Res 2017 May 26;19(5):e147 [FREE Full text] [doi: 10.2196/jmir.7463] [Medline: 28550001]
34. Bargh J. Automaticity in action: the unconscious as repository of chronic goals and motives. In: Gollwitzer PM, Bargh JA, editors. The Psychology of Action: Linking Cognition and Motivation to Behavior. New York: Guilford Press; 1996:457-481.
35. Nicolas G, Skinner AL. "That's so gay!" Priming the general negative usage of the word gay increases implicit anti-gay bias. J Soc Psychol 2012;152(5):654-658. [doi: 10.1080/00224545.2012.661803] [Medline: 22931002]
36. Williams LE, Bargh JA. Experiencing physical warmth promotes interpersonal warmth. Science 2008 Oct 24;322(5901):606-607 [FREE Full text] [doi: 10.1126/science.1162548] [Medline: 18948544]
37. Kelly JF, Westerhoff CM. Does it matter how we refer to individuals with substance-related conditions? a randomized study of two commonly used terms. Int J Drug Policy 2010 May;21(3):202-207. [doi: 10.1016/j.drugpo.2009.10.010] [Medline: 20005692]
38. Nahum-Shani I, Smith S, Spring B, Collins L, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
39. Gollwitzer PM. Implementation intentions: strong effects of simple plans. Am Psychol 1999;54(7):493-503. [doi: 10.1037/0003-066x.54.7.493]
40. Tackett-Newburg K. Choose the path of least friction to change your behavior. Psychology Today. 2020 Jan 19. URL: https://www.psychologytoday.com/ca/blog/mentality-check/202001/choose-the-path-least-friction-change-your-behavior [accessed 2020-05-14]
41. Clark R, Mayer R. Scenario-Based e-Learning: Evidence-Based Guidelines for Online Workforce Learning. San Fransisco, California: John Wiley & Sons; 2012.
42. Berridge KC, Robinson TE. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Res Rev 1998 Dec;28(3):309-369. [doi: 10.1016/s0165-0173(98)00019-8]
43. Bouton M. Learning and Behavior: A Contemporary Synthesis. Sunderland, Massachusetts: Sinauer Associates; 2007.
44. Fogg B. Tiny Habits: The Small Changes That Change Everything. Boston: Houghton Mifflin Harcourt; 2019.
45. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4.
46. Stock J, Cervone D. Proximal goal-setting and self-regulatory processes. Cogn Ther Res 1990 Oct;14(5):483-498. [doi: 10.1007/bf01172969]
47. Skinner B. Science and Human Behavior. Oxford, England: Macmillan; 1965.
48. Fleckman JM, Taylor CA, Storer HL, Andrinopoulos K, Weil LE, Rubin-Miller L, et al. Breaking the mold: socio-ecologic factors to influence the development of non-harsh parenting strategies to reduce risk for child physical abuse. Child Youth Serv Rev 2018 Nov;94:274-283. [doi: 10.1016/j.childyouth.2018.10.019]
49. Dede Yildirim E, Roopnarine JL. Positive discipline, harsh physical discipline, physical discipline and psychological aggression in five Caribbean countries: associations with preschoolers' early literacy skills. Int J Psychol 2019 Jun 02;54(3):342-350. [doi: 10.1002/ijop.12465] [Medline: 29094762]
50. Treadway MT, Buckholtz JW, Schwartzman AN, Lambert WE, Zald DH. Worth the 'EEfRT'? the effort expenditure for rewards task as an objective measure of motivation and anhedonia. PLoS One 2009 Aug 12;4(8):e6598 [FREE Full text] [doi: 10.1371/journal.pone.0006598] [Medline: 19672310]

51. Muench F, van Stolk-Cooke K, Kuerbis A, Stadler G, Baumel A, Shao S, et al. A randomized controlled pilot trial of different mobile messaging interventions for problem drinking compared to weekly drink tracking. PLoS One 2017 Feb 1;12(2):e0167900 [FREE Full text] [doi: 10.1371/journal.pone.0167900] [Medline: 28146560]
52. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82.
53. Baumel A, Birnbaum ML, Sucala M. A systematic review and taxonomy of published quality criteria related to the evaluation of user-facing eHealth programs. J Med Syst 2017 Aug;41(8):128. [doi: 10.1007/s10916-017-0776-6] [Medline: 28735372]
54. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding health behavior technology engagement: pathway to measuring digital behavior change interventions. JMIR Form Res 2019 Oct 10;3(4):e14052. [doi: 10.2196/14052] [Medline: 31603427]
55. Baumel A, Muench F. Heuristic evaluation of ehealth interventions: establishing standards that relate to the therapeutic process perspective. JMIR Ment Health 2016 Jan 13;3(1):e5 [FREE Full text] [doi: 10.2196/mental.4563] [Medline: 26764209]
56. Cummins CO, Prochaska JO, Driskell M, Evers KE, Wright JA, Prochaska JM, et al. Development of review criteria to evaluate health behavior change websites. J Health Psychol 2003 Jan;8(1):55-62. [doi: 10.1177/1359105303008001434] [Medline: 22113900]
57. Monkman H, Kushniruk A. A health literacy and usability heuristic evaluation of a mobile consumer health application. Stud Health Technol Inform 2013;192:724-728. [Medline: 23920652]
58. Baumel A, Correll CU, Birnbaum M. Adaptation of a peer based online emotional support program as an adjunct to treatment for people with schizophrenia-spectrum disorders. Internet Interv 2016 May;4:35-42 [FREE Full text] [doi: 10.1016/j.invent.2016.03.003] [Medline: 30135788]
59. Harari Y. Homo Deus: A Brief History of Tomorrow. London: Harvill Secker; 2016.
60. Festinger L. A Theory of Cognitive Dissonance Vol 2. Redwood City, California: Stanford University Press; 1957.
61. Shiv B, Carmon Z, Ariely D. Placebo effects of marketing actions: consumers may get what they pay for. J Mark Res 2018 Oct 16;42(4):383-393. [doi: 10.1509/jmkr.2005.42.4.383]
62. Ariely D. Payoff: The Hidden Logic That Shapes Our Motivations. New York, NY: Simon and Schuster; 2016.
63. de Saint-Exupe´ry A. The Little Prince. New York: Reynal & Hitchcock; 1943.
64. Eyal N, Hoover R. Hooked: How to Build Habit-Forming Products. New York: Penguin Group; 2014.
65. Webster-Stratton C. The Incredible Years: A Trouble-Shooting Guide for Parents of Children Aged 3-8 Years. Seattle, Washington: Umbrella Press; 2003.
66. McMahon R, Long N, Forehand R. Parent training for the treatment of oppositional behavior in young children: helping the noncompliant child. In: Murrihy RC, Kidman AD, Ollendick TH, editors. Clinical Handbook of Assessing and Treating Conduct Problems in Youth. New York: Springer; 2011:163-191.
67. Sanders MR. Triple P-Positive Parenting Program: towards an empirically validated multilevel parenting and family support strategy for the prevention of behavior and emotional problems in children. Clin Child Fam Psychol Rev 1999 Jun;2(2):71-90. [doi: 10.1023/a:1021843613840] [Medline: 11225933]
68. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
69. Kientz J, Choe E, Birch B, Maharaj R, Fonville A, Glasson C, et al. Heuristic evaluation of persuasive health technologies. In: Proceedings of the 1st ACM International Health Informatics Symposium. 2010 Presented at: 1st ACM International Health Informatics Symposium; November 11-12; Arlington, Virginia. [doi: 10.1145/1882992.1883084]
70. Holter MTS, Johansen A, Brendryen H. How a fully automated eHealth program simulates three therapeutic processes: a case study. J Med Internet Res 2016 Jun 28;18(6):e176 [FREE Full text] [doi: 10.2196/jmir.5415] [Medline: 27354373]
71. Banks HT, Rehm KL, Sutton KL, Davis C, Hail L, Kuerbis A, et al. Dynamic modeling of behavior change. Quart Appl Math 2014 Apr 1;72(2):209-251. [doi: 10.1090/s0033-569x-2014-01296-3]
72. Luborsky L, Rosenthal R, Diguer L, Andrusyna T, Berman J, Levitt J, et al. The dodo bird verdict is alive and well—mostly. Clin Psychol 2002;9(1):2-12. [doi: 10.1093/clipsy.9.1.2]

Edited by G Eysenbach; submitted 09.10.20; peer-reviewed by M Torok, N Khalili-Mahani; comments to author 02.11.20; revised version received 16.12.20; accepted 16.01.21; published 12.03.21.

Please cite as:
Baumel A, Muench FJ
Effort-Optimized Intervention Model: Framework for Building and Analyzing Digital Interventions That Require Minimal Effort for Health-Related Gains
J Med Internet Res 2021;23(3):e24905
URL: https://www.jmir.org/2021/3/e24905
doi:10.2196/24905
PMID:33709943

Original Paper

The Reliability of Remote Patient-Reported Outcome Measures via Mobile Apps to Replace Outpatient Visits After Rotator Cuff Repair Surgery: Repetitive Test-Retest Comparison Study for 1-Year Follow-up

Taek Ho Hong[1], MD; Myung Ku Kim[1], MD, PhD; Dong Jin Ryu[1], MD; Jun Sung Park[1], MD; Gi Cheol Bae[1], MD; Yoon Sang Jeon[1], MD, PhD

Department of Orthopedic Surgery, Inha University Hospital, Incheon, Republic of Korea

Corresponding Author:
Yoon Sang Jeon, MD, PhD
Department of Orthopedic Surgery
Inha University Hospital
27 Inhang-ro
Jung-gu
Incheon, 22332
Republic of Korea
Phone: 82 010 8353 3695
Email: ysjeon80@hanmail.net

Abstract

Background: With the development of health care–related mobile apps, attempts have been made to implement remote patient-reported outcome measures (PROMs). In order for remote PROMs to be widely used by mobile apps, the results should not be different depending on the location; that is, remote PROM results performed in locations other than hospitals should be able to obtain reliable results equivalent to those performed in hospitals, and this is very important. However, to our knowledge, there are no studies that have assessed the reliability of PROMs using mobile apps according to the location by comparing the results performed remotely from the hospital and performed at the outpatient visits.

Objective: The purpose of this study was to evaluate the reliability of remote PROMs using mobile apps compared to PROMs performed during outpatient follow-up visits after arthroscopic shoulder surgery.

Methods: A total of 174 patients who underwent arthroscopic rotator cuff repair completed questionnaires 2 days before visiting the clinic for the 1-, 2-, 3-, 6-, and 12-month follow-ups (test A). The patients completed the questionnaires at the clinic (test B) using the same mobile app and device for the 1-, 2-, 3-, 6-, and 12-month follow-ups. Test-retest comparisons were performed to analyze the differences and reliability of the PROMs according to the period.

Results: Comparisons of tests A and B showed statistically significant differences at 1, 2, and 3 months (all *P*s<.05 except for the ASES function scale at 3-months) but not 6 or 12 months after surgery (all *P*s>.05). The intraclass correlation values between the two groups were relatively low at the 1-, 2-, and 3-month follow-ups but were within the reliable range at 6 and 12 months after surgery. The rate of completion of tests A and B using the mobile app was significantly lower in the group older than 70 years than in the other groups for all postoperative periods (*P*<.001).

Conclusions: PROMs using mobile apps with different locations differed soon after surgery but were reliably similar after 6 months. The remote PROMs using mobile apps could be used reliably for the patient more than 6 months after surgery. However, it is to be expected that the use of mobile app–based questionnaires is not as useful in the group older than 70 years as in other age groups.

(J Med Internet Res 2021;23(3):e20989) doi:10.2196/20989

KEYWORDS

patient-reported outcome measures (PROMs); location; remote PROMs using mobile application; smartphone; mobile phone; follow-up loss

Introduction

Background

With the dramatically increased penetration rates worldwide [1], at 81% in the United States and 95% in South Korea [2], smartphones are becoming increasingly indispensable in everyday life [3]. A variety of mobile apps for information, communication, education, and entertainment purposes have been developed for smartphones [3], including mobile health care systems. Seto et al [4] developed a mobile phone–based telemonitoring program for patients with heart failure following acute decompensation. Denono et al [5] suggested that postoperative mobile apps after ambulatory lumbar discectomy were effective tools for spine surgeons.

With the development of health care–related mobile apps, attempts have been made to implement remote patient-reported outcome measures (PROMs). Skrepnik et al [6] assessed the impact of a novel smartphone app compared with standard follow-up on mobility following treatment with intra-articular injection in patients with knee osteoarthritis. Armstrong et al [7] evaluated the effect of home monitoring via a mobile app on the number of in-person visits following ambulatory surgery. Most studies reported that patients found mobile apps for remote follow-ups to be convenient, safe, and highly satisfactory [4-8]. Reliable remote follow-ups by mobile health care systems have several advantages over face-to-face follow-ups. In general, follow-up durations of at least 12 months to several years are required for reliable clinical study findings after surgery [9,10]. However, maintaining high rates of long-term follow-up is challenging due to poor patient compliance [10,11]. Remote follow-ups using mobile PROMs are also efficient in terms of health care costs compared to outpatient visits [12]. Considering the difficulty in long-term follow-up [10], the reduction in outpatient follow-ups, and the reduced health care costs [12], PROMs using mobile apps performed outside of clinics may be good alternatives. In order for remote PROMs to be widely used by the mobile app, the results should not be different depending on the location; that is, remote PROM results performed in locations other than hospitals should be able to obtain reliable results equivalent to those performed in hospitals, and this is very important. However, to our knowledge, there are no studies that have assessed the reliability of PROMs using mobile apps according to the location by comparing the results performed remotely from the hospital and performed at the outpatient visits.

Goal of This Study

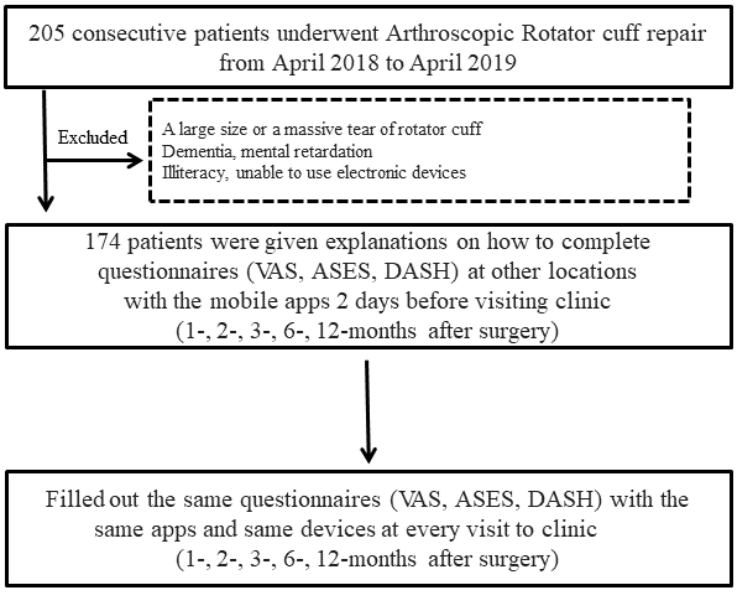
Therefore, this study evaluated the reliability of remote PROMs using mobile apps compared to the PROMs performed by the same mobile apps during outpatient follow-up visits after arthroscopic shoulder surgery. We also analyzed the tendencies in differences with increasing time after surgery and observed the PROM participation rates of patients according to the follow-up periods with repetitive test-retest studies. We hypothesized that the results of the PROMs would be similar between those measured in outpatient clinic visits and those measured remotely using mobile apps.

Methods

Patients and Study Design

205 consecutive patients who underwent arthroscopic rotator cuff repair by a single surgeon were initially considered for this study between April 2018 and April 2019. Patients diagnosed with large or massive rotator cuff tears were excluded because of the difference in their rehabilitation schedules. Patients with dementia, mental retardation, illiteracy, or inability to use electronic devices were excluded because of the difficulty in completing questionnaires using electronic equipment. After exclusion, the remaining 174 patients (92 men and 82 women) prospectively conducted the test-retest comparisons, which were performed 5 times each to assess the results after surgery. The patients were instructed to complete questionnaires (visual analog scale [VAS], American Shoulder and Elbow Society [ASES] scale [13], and Disabilities of the Arm, Shoulder, and Hand [DASH] scale [14]) at other locations (test A) 2 days before visiting the clinic for the 1-, 2-, 3-, 6-, and 12-month postoperative follow-ups. Using the same app and electronic devices, namely, their mobile phones, each patient completed the same questionnaires at the clinic (test B) at 1, 2, 3, 6, and 12 months after surgery (Figure 1). The patients received mobile messages linked to an app for an electronic PROM system (Proscore, Incheon, South Korea). All patients who visited our clinic answered the same questionnaires with the mobile app before treatment. The timing of mobile messaging was determined to be 48 hours before the clinic visit based on a previous systematic review that reported test-retest reliability [15]. Of the 174 patients, test A (PROMs completed via the mobile app installed on the mobile phone of each patient 2 days before the clinic visit) was completed by 148 at 1 month, 135 at 2 months, 106 at 3 months, 77 at 6 months, and 59 at 12 months. All 174 patients visited our clinic at 1 month after surgery. However, the rates of outpatient visits with patients completing test B using the same app and electronic devices (the mobile phone of each patient) decreased over time, with 170 visiting at 2 months, 142 at 3 months, 112 at 6 months, and 95 at 12 months after surgery (Figure 1). All patients underwent the same course of rehabilitation. An abduction brace was applied for 4 weeks after surgery. Passive range of motion exercises were allowed from 4 to 8 weeks after surgery. Active range of motion exercises were conducted 8 weeks after surgery. This study, including the subject selection and data collection, was conducted under the approval of the Inha University Hospital Institutional Review Board (IRB INHA 2019-09-024) in accordance with the 1964 Declaration of Helsinki.

Figure 1. Inclusion and exclusion criteria for the study. ASES: American Shoulder and Elbow Society; DASH: Disabilities of the Arm, Shoulder, and Hand; VAS: visual analog scale.



Scale Definitions and Measures

The VAS score is measured ranging from 0 to 10, with scores of 0 and 10 indicating “no pain” and “worst pain imaginable,” respectively. The ASES scale [13] consists of two subscales, namely, pain (1 item) and function (10 items). Each subscale is transformed to scores ranging from 0 to 50, based on patient responses. The sum of the two scales is the total score on the ASES scale, with a score of 100 points indicating perfect conditions of the shoulder. This study analyzed the total ASES scale score as well as the scores for the two subscales. The DASH scale comprises 30 items (21 on daily activities, 5 on symptoms, 3 on participation, and 1 on confidence in ability) [14]. Higher scores indicate worse upper limb function. We used an electronic PROM system (Proscore, Incheon, South Korea) available as an app for electronic devices that measures VAS, ASES scale, and DASH scale scores at locations other than the clinic. In this system, patients touched the answer on the screen instead of marking their responses on original paper questionnaires using a writing instrument. This change from paper-based to electronic-based measures is minor, according to the Food and Drug Administration guidelines [16].

Statistical Analyses

The data are expressed as means (standard deviations) or medians (ranges). Paired *t* tests (2-tailed) were used to evaluate differences between the answers for tests A and B; more specifically, the average score with standard deviations of the scale’s scores was calculated and analyzed using paired *t* tests. We also calculated the average absolute value of the differences between tests A and B. Intraclass correlation coefficients (ICCs) were calculated to estimate reproducibility and reliability between tests A and B. Statistical significance was indicated by $P<.05$. All statistical analyses were performed using IBM SPSS Statistics for Windows, version 19.0 (IBM Corp, Armonk, NY).

Results

The demographics of patients undergoing rotator cuff surgery are summarized in Table 1.

The average scores and absolute values of the differences between tests A and B are shown in Table 2 and Figure 2 for the 1-, 2-, 3-, 6-, and 12-month postoperative results. At 1, 2, and 3 months after surgery, test B showed significantly better outcomes compared to those of test A (Ps<.05), except for the ASES function subscale (P=.06 at 3 months). All parameters did not show statistically significant differences (all Ps>.05) between tests A and B at 6 and 12 months after surgery. The average absolute differences in VAS, ASES total, and DASH scores between tests A and B were 1.68, 14.72 and 11.28 points at 1 month after surgery, respectively. In most of the scales, the differences in the average and absolute differences gradually decreased with time after surgery. At 12 months after surgery, the average absolute value differences in VAS, ASES total, and DASH scores between tests A and B were greatly reduced (0.32, 5.48, and 4.46 points, respectively).

Table 1. Baseline demographic and clinical characteristics (N=174).

Characteristic	Value
Age (years), mean (SD)	59.38 (10.9)
Gender, female, n (%)	82 (47.1)
Side, n (%)	
Right	97 (55.7)
Left	77 (44.3)
Symptom duration (months), mean (SD)	11.18 (13.74)
Tear size, n (%)	
Small	96 (55.2)
Medium	78 (44.8)

Table 2. Mean (standard deviation) for each scale by 1-, 2-, 3-, 6-, and 12-month postoperative data analyzed by paired *t* test or Wilcoxon signed rank test (N=174).

POD[a] and scale	Test A, mean (SD)	Test B, mean (SD)	Differences	*P* value	Absolute differences[b]
POD 1 month					
VAS[c] score	3.23 (1.69)	1.96 (1.16)	1.27	<.001	1.68 (1.23)
ASES[d] total	50.02 (11.78)	60.56 (14.85)	−10.54	<.001	14.72 (10.07)
ASES pain	29.42 (8.85)	34.32 (11.73)	−4.90	<.001	11.28 (7.53)
ASES function	20.60 (6.40)	26.24 (7.51)	−5.64	<.001	7.53 (6.26)
DASH[e]	64.01 (10.57)	53.82 (11.92)	10.19	<.001	12.94 (9.35)
POD 2 months					
VAS score	2.44 (1.62)	1.60 (1.17)	0.84	<.001	1.25 (1.04)
ASES total	54.88 (15.93)	63.68 (12.89)	−8.80	<.001	13.85 (9.41)
ASES pain	32.55 (12.85)	35.96 (9.35)	−3.41	.003	10.88 (8.28)
ASES function	22.33 (7.62)	27.72 (8.52)	−5.39	<.001	8.40 (6.46)
DASH	54.26 (9.82)	48.34 (11.81)	5.92	<.001	9.49 (6.93)
POD 3 months					
VAS score	2.24 (1.64)	1.54 (1.05)	0.70	.03	1.28 (0.95)
ASES total	62.73 (12.05)	67.15 (11.62)	−4.42	.01	11.43 (9.10)
ASES pain	34.57 (9.93)	36.88 (8.00)	−2.31	.02	7.87 (6.36)
ASES function	28.16 (6.82)	30.27 (7.37)	−2.11	.06	6.60 (5.35)
DASH	48.10 (9.26)	43.36 (12.32)	4.74	<.001	8.86 (7.06)
POD 6 months					
VAS score	1.06 (0.54)	0.88 (0.70)	0.18	.13	0.46 (0.59)
ASES total	74.28 (12.15)	77.19 (13.20)	−2.91	.09	8.78 (7.09)
ASES pain	37.07 (8.86)	39.35 (7.83)	−2.28	.21	5.90 (5.94)
ASES function	37.20 (7.82)	37.84 (9.69)	−0.64	.30	5.26 (4.82)
DASH	36.64 (10.95)	33.10 (9.54)	3.54	.48	7.73 (4.94)
POD 12 months					
VAS score	0.83 (0.56)	0.77 (0.58)	0.06	.49	0.32 (0.47)
ASES total	78.78 (9.07)	79.96 (10.94)	−1.18	.17	5.48 (3.76)
ASES pain	41.69 (6.53)	42.45 (5.67)	−0.76	.24	3.81 (3.75)
ASES function	37.08 (7.32)	37.51 (8.87)	−0.43	.32	3.47 (2.80)
DASH	31.34 (8.81)	30.29 (7.66)	1.05	.19	4.46 (3.74)

[a]POD: postoperative duration.

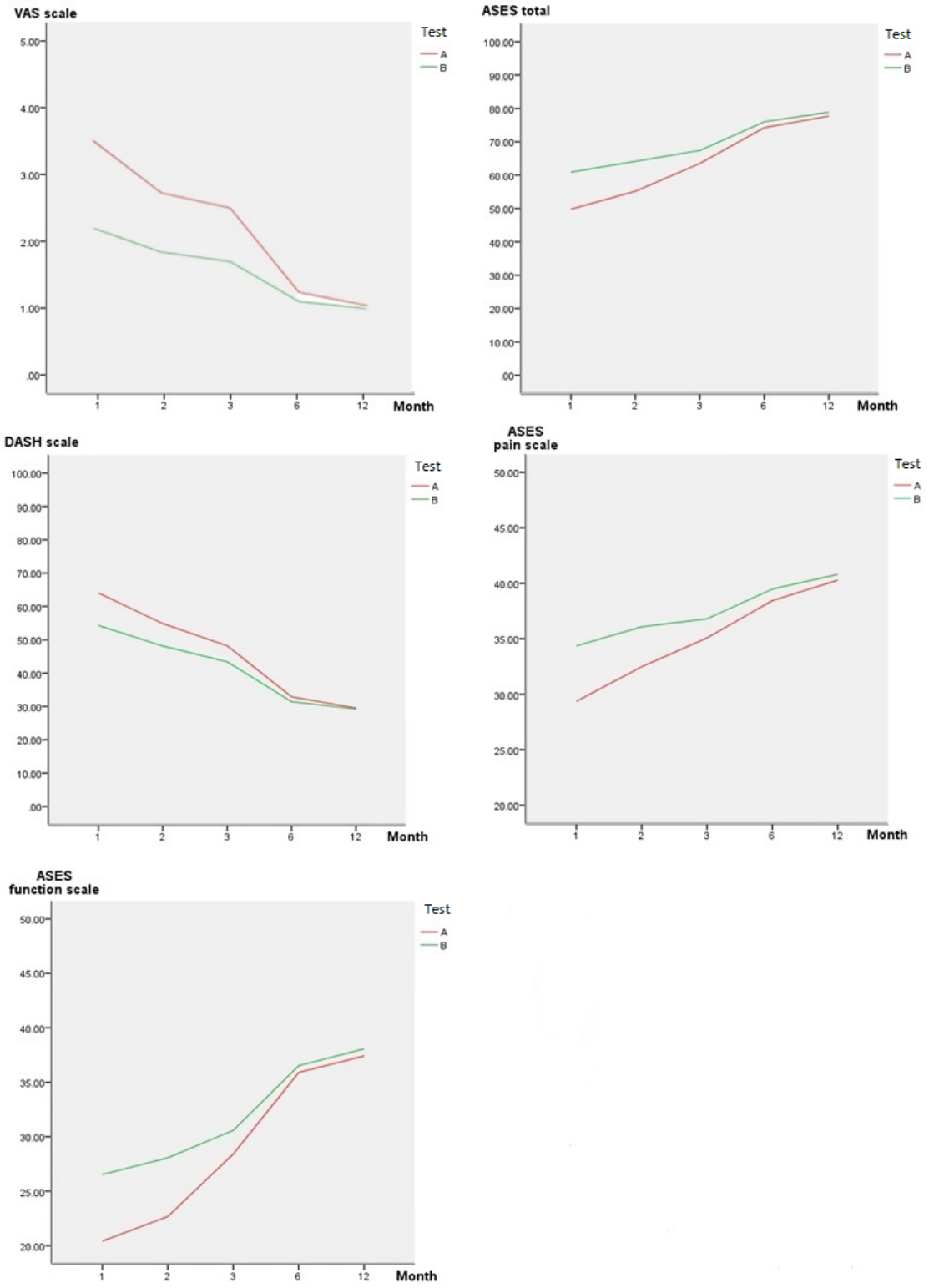
[b]Absolute differences are calculated by taking the greater value minus the smaller one between tests A and B.

[c]VAS: visual analog scale.

[d]ASES: American Shoulder and Elbow Society Shoulder Index.

[e]DASH: Disabilities of the Arm, Shoulder, and Hand score.

Figure 2. Mean (standard deviation) for each scale used in this study. Data presented for 5 tests by postoperative duration (N=174). ASES: American Shoulder and Elbow Society; DASH: Disabilities of the Arm, Shoulder, and Hand; VAS: visual analog scale.



To estimate the reproducibility and reliability between tests A and B, ICC values were calculated for each scale and subscale (Table 3). The VAS scale and ASES pain subscale showed relatively low ICC values compared to those of the other scales. The lowest ICC value for the VAS scale was observed at 1 month after surgery (0.51, moderate reliability). The low ICC values for the ASES pain subscale were observed at 1, 2, and 3 months after surgery (0.47, 0.46, and 0.47, respectively; poor reliability). Moderate ICC values were observed for the ASES function subscale at 1, 2, and 3 months after surgery (0.50, 0.53, and 0.67, respectively). At 6 months after surgery, all parameters showed good ICC values (0.77 for VAS, 0.83 for DASH scale, 0.80 for ASES function subscale, 0.78 for ASES pain subscale, and 0.78 for ASES total scale). Regarding the DASH scale, a good ICC value was observed at 6 months after surgery (0.83). The highest ICC values for all parameters were observed at 12

months after surgery. VAS score, ASES pain subscale, and DASH scale showed good ICC values at 12 months after surgery (0.81, 0.76, and 0.87, respectively). The ASES function scale and the ASES total scale showed excellent ICC values at 12 months after surgery (0.91 and 0.90, respectively).

Table 3. Intraclass correlation coefficient values for each scale (N=174).

Scale/subscale	Postoperative duration (months)				
	1	2	3	6	12
VAS[a] pain	0.51	0.67	0.62	0.77	0.81
ASES[b]					
Pain	0.47	0.46	0.47	0.78	0.76
Function	0.54	0.65	0.67	0.8	0.91
Total	0.5	0.53	0.58	0.78	0.9
DASH[c] total	0.57	0.72	0.71	0.83	0.87

[a]VAS: visual analog scale.

[b]ASES: American Shoulder and Elbow Society Shoulder Index.

[c]DASH: Disabilities of the Arm, Shoulder and Hand score.

The rates of outpatient visits and completions of tests A and B according to the period for each age group are shown in Table 4. At 1 month, all 174 patients visited our clinic. However, with time after surgery, the number of outpatient visits gradually decreased. No significant differences in the numbers of outpatient visits were observed in terms of age (*P*=.60, .54, .91, and .70 for 2, 3, 6, and 12 months after surgery, respectively). The rate of completion of tests A and B using the mobile app was significantly lower in the group older than 70 years than in the other groups for all postoperative periods (*P*<.001).

Table 4. Visit rate and test-retest response rate by age at each postoperative period.

Rates by postoperative period (months)	Age (years), n (%)					*P* value
	<50 (n=33)	50-59 (n=51)	60-69 (n=52)	≥70 (n=38)	Total (N=174)	
Outpatient visits (visit rate)						
1	33 (100)	51 (100)	52 (100)	38 (100)	174 (100)	—
2	32 (97.0)	51 (100)	50 (96.2)	37 (97.4)	170 (97.7)	.60
3	28 (84.8)	42 (82.4)	44 (84.6)	28 (73.7)	142 (81.6)	.54
6	20 (60.6)	32 (62.7)	34 (65.4)	26 (68.4)	112 (64.4)	.91
12	15 (45.5)	29 (56.9)	30 (57.7)	21 (55.3)	95 (54.6)	.70
Test-retest responses (response/visit rate)						
1	32 (97.0)	48 (94.1)	47 (90.4)	21 (55.3)	148 (85.1)	<.001
2	30 (93.7)	45 (88.2)	43 (86.0)	17 (45.9)	135 (79.4)	<.001
3	25 (89.3)	37 (88.1)	36 (81.8)	8 (28.6)	106 (74.6)	<.001
6	19 (95.0)	26 (81.2)	25 (73.5)	7 (26.9)	77 (68.7)	<.001
12	13 (86.7)	22 (75.9)	19 (63.3)	5 (23.8)	59 (62.1)	<.001

In this study, 36 of 174 patients (20.7%) completed all follow-up visits (1, 2, 3, 6, and 12 months after surgery) and also completed tests A and B (completely implemented group). We also performed comparisons between tests A and B in this group to determine the average difference for each scale (Table 5). The ICC values between tests A and B in the completely implemented group (n=36) were similar to those for all 174 patients (Table 6). At 1, 2, and 3 months after surgery, test B showed significantly better outcomes than those of test A (*P*s<.05), except for the ASES pain subscale and DASH scale. No parameter differed significantly between tests A and B at 6 and 12 months after surgery. The average absolute value of the differences for the VAS, ASES total, and DASH scores between tests A and B were 1.50, 15.97, and 10.28 points, respectively, at 1 month after surgery. In most of the scales, the average and the absolute differences gradually decreased with time after surgery. All parameters showed poor or moderate ICCs at 1, 2, and 3 months after surgery but showed moderate or good values at 6 months and peaked at 12 months after surgery for all parameters. The VAS score, ASES pain subscale, and ASES total scale showed good ICCs at 12 months after surgery (0.80, 0.82, and 0.88 respectively), while the ASES function scale and

the DASH scale showed excellent ICCs at 12 months after surgery (0.92 and 0.90, respectively).

Table 5. Mean (standard deviation) for each scale by 1-, 2-, 3-, 6-, and 12-month postoperative data analyzed by paired *t* test or Wilcoxon signed rank test in the completely implemented group (N=36).

POD[a] and scale	Test A, mean (SD)	Test B, mean SD)	Differences	*P* value	Absolute differences[b]
POD 1 month					
VAS[c] score	3.05 (1.58)	1.88 (1.06)	1.17	.009	1.50 (1.02)
ASES[d] total	51.24 (10.84)	63.19 (13.11)	−11.95	<.001	15.97 (9.92)
ASES pain	30.27 (9.01)	35.97 (10.33)	−5.70	.02	10.28 (7.35)
ASES function	20.96 (7.18)	27.21 (7.23)	−6.25	<.001	8.10 (5.97)
DASH[e]	64.75 (12.68)	55.48 (11.39)	9.27	<.001	13.41 (8.86)
POD 2 months					
VAS score	2.41 (1.61)	1.55 (1.22)	0.86	.004	1.47 (1.05)
ASES total	56.94 (12.20)	66.75 (9.08)	−9.81	<.001	14.81 (9.99)
ASES pain	33.33 (13.09)	37.78 (8.49)	−3.45	.03	10.01 (7.36)
ASES function	23.60 (6.75)	28.97 (7.42)	−5.37	<.001	8.14 (5.26)
DASH	54.20 (9.82)	47.19 (10.20)	7.01	<.001	9.54 (9.15)
POD 3 months					
VAS score	2.24 (1.64)	1.54 (1.05	0.70	.01	1.23 (1.09)
ASES total	62.51 (12.07)	67.15 (11.62)	−4.64	.02	12.73 (9.82)
ASES pain	34.57 (9.93)	36.88 (8.00)	−2.31	.07	7.77 (6.80)
ASES function	27.93 (6.65)	30.26 (7.37)	−2.33	.03	7.54 (5.51)
DASH	48.10 (9.26)	43.36 (12.32)	4.74	.18	8.85 (6.81)
POD 6 months					
VAS score	1.06 (0.54)	0.88 (0.70)	0.18	.12	0.53 (0.60)
ASES total	74.28 (12.15)	77.19 (13.20)	−2.91	.18	9.16 (7.17)
ASES pain	37.07 (8.86)	39.35 (7.83)	−2.28	.09	5.69 (6.67)
ASES function	37.20 (7.82)	37.84 (9.69)	−0.64	.73	5.51 (5.09)
DASH	36.64 (10.95)	33.10 (9.54)	3.54	.28	6.49 (4.36)
POD 12 months					
VAS score	0.83 (0.56)	0.77 (0.58)	0.06	.17	0.36 (0.48)
ASES total	78.78 (9.07)	79.96 (10.94)	−1.18	.16	4.67 (3.39)
ASES pain	41.69 (6.53)	42.45 (5.67)	−0.76	.15	3.61 (3.50)
ASES function	37.08 (7.32)	37.51 (8.87)	−0.43	.56	3.38 (3.07)
DASH	31.34 (8.81)	30.29 (7.66)	1.05	.41	3.61 (2.52)

[a]POD: postoperative duration.

[b]Absolute differences are calculated by taking the greater value minus the smaller one, between tests A and B.

[c]VAS: visual analog scale.

[d]ASES: American Shoulder and Elbow Society Shoulder Index.

[e]DASH: Disabilities of the Arm, Shoulder, and Hand score.

Table 6. Intraclass correlation coefficient values for the completely implemented group (N=36).

Scale/subscale	Postoperative duration (months)				
	1	2	3	6	12
VAS[a] pain	0.45	0.54	0.53	0.76	0.80
ASES[b]					
Pain	0.53	0.56	0.54	0.78	0.82
Function	0.56	0.5	0.58	0.81	0.92
Total	0.56	0.52	0.57	0.81	0.88
DASH[c] total	0.56	0.60	0.53	0.84	0.90

[a]VAS: visual analog scale.

[b]ASES: American Shoulder and Elbow Society Shoulder Index.

[c]DASH: Disabilities of the Arm, Shoulder and Hand score.

Discussion

The results of this study revealed that the PROMs varied depending on the location for the initial 1-, 2-, and 3-month follow-ups after arthroscopic shoulder surgery. However, at 6 months or more after surgery, the PROMs using the mobile apps showed similar results regardless of location. The ICC analysis also showed a tendency toward relatively low values for 1, 2, and 3 months postoperatively according to the PROM location, while high values were recorded at the 6- and 12-month follow-ups. These findings indicated that PROMs performed using mobile apps at 6 months after surgery were adequately reliable and reproducible regardless of location. Therefore, the use of remote PROMs via mobile apps may be more valuable for follow-ups at 6 months or more after surgery, when the rate of follow-up loss is increased.

Most scales showed different outcomes for test B compared to those for test A at the initial 1, 2, and 3 months postsurgery. However, at 6 and 12 months after surgery, none of the scales differed significantly between tests A and B. The absolute values of the differences were also greatly reduced with time, and the reliability as assessed by ICC was adequately high after 6 months. These outcomes are consistent with those of previous studies on the test-retest reliability of PROMs. Chahal et al [17] reported good reliability of PROM for knee joint–specific questionnaires in a test-retest study conducted 6 months after multiligament knee injury. Bramming et al [18] reported that a PROM (forgotten joint score-12) showed high relative reliability in a test-retest study conducted at 6 months after hip arthroscopic surgery. The differences in follow-ups performed in the first 3 months postsurgery might be due to variability in patient conditions during the acute phase following surgery. Additionally, the differences may have decreased over time due to patients getting used to the test items by repeatedly performing PROMs. The absolute values of the differences between the two tests were also noteworthy, given that the purpose of this study was to measure the difference between outpatient and remote mobile apps. The absolute values of the differences for each scale were relatively high at the 1-, 2-, and 3-month follow-ups. However, at the 6- and 12-month follow-ups, all parameters showed reduced absolute differences. These results also reinforce the reliability of the remote PROMs compared to outpatient PROMs for long-term follow-ups.

Clinical studies on patient outcomes after surgery generally require at least 12 months to several years of follow-up for recognition as reliable clinical studies [9,10]. To avoid biases in clinical studies using PROMs performed at the clinic, it is important to minimize loss to follow-up to the hospital [19,20]. However, maintaining high rates of long-term follow-up is challenging due to poor patient compliance [10]. Cronin et al [21] showed that 40% of patients with orthopedic trauma did not complete 90 days of follow-up. Zelle et al [19] also reported that patients with undifferentiated orthopedic trauma showed high rates (>70%) of noncompliance in the initial 6 months postsurgery. Considering that patients' compliance with outpatient follow-up decreases over time after surgery [21], the reliability of PROMs via mobile apps regardless of the location for long-term follow-up after surgery is meaningful as these PROMs may be an option to assess patient condition without a need to travel to the hospital.

Even in terms of the cost benefits and efficient follow-ups for patients [12], the reliability of remote PROMs is also important. Higgins et al [22] compared a conventional in-person visit follow-up group (conventional group) to a non–face-to-face follow-up group using a mobile app (mobile app group) for 6 weeks after anterior cruciate ligament reconstruction. The mobile app group had 0.36 clinic visits during the study period, compared to 2.44 visits in the conventional group. The mobile app group spent Can $211 (US $166.16) less over 6 weeks than the other group. Thus, in terms of cost burden, remote PROMs may also have advantages over outpatient visits if the assessments are reliable.

Due to the recent infectious disease epidemics of COVID-19 [11], it is difficult to expect patients to comply with outpatient follow-ups in the absence of an emergency [23]. Remote PROMs are particularly valuable [11] as medical staff and national health care system resources are focused on a particular infectious disease [24-26]. Recent guidelines from the Journal of Bone and Joint Surgery [27] recommend the assessment of all planned elective or nonemergency surgical procedures and clinical visits to determine whether they can be postponed or canceled. If remote PROMs are reliable, they can be effective

and highly utilized for reducing patient visits [11] and allow efficient distribution of the national health care system capacity when infectious disease outbreaks occur.

This study used an electronic PROM system (Proscore, Incheon, South Korea) available for mobile phones. The correlation between electronic measuring systems and conventional paper-and-pencil methods is reportedly reliable [28]. The compliance of patients for completing scoring tools using electronic systems is generally better than that for paper-and-pencil methods because it is more convenient and quicker [29]. However, older patients may not prefer performing PROMs with electronic devices because of less exposure to and familiarity with electronic devices compared to younger patients [29]. In this study, the rate of outpatient visits did not differ significantly by age; however, the rates of test-retest completion for both PROMs at outpatient visits and remote PROMs using mobile apps were statistically significantly lower in patients older than 70 years than those in other groups for all postoperative periods (P<.001). Instructions for the use of smartphone devices and apps must be provided to the elderly in order to use PROMs via mobile apps at locations other than hospitals.

Test-retest assessments to evaluate the reliability of tools are generally conducted once for comparisons. However, this study conducted test-retest comparisons 5 times each to determine the tendencies with increasing time after surgery, which is a strength of this study. The limitations of this study were its inclusion of only patients who underwent arthroscopic rotator cuff repair. However, this could also be considered a strength as confounding variables due to many disease entities are reduced. Several diseases and treatment options for the shoulder joint, including intra-articular injection for frozen shoulder, reverse total shoulder arthroplasty for rotator cuff arthropathy, and other disease categories, might be candidates for further study.

In conclusion, PROMs performed using mobile apps in different locations showed varied results soon after surgery but were similar after 6 months, with reliable ICC values. The remote PROMs using mobile apps could be used reliably for the patient more than 6 months after surgery. However, it is to be expected that the use of mobile app–based questionnaires is not as useful in the group older than 70 years as in other age groups.

Acknowledgments

This paper was supported by Inha University.

Conflicts of Interest

None declared.

References

1. James J. Penetration and Growth Rates of Mobile Phones in Developing Countries: An Analytical Classification. Soc Indic Res 2010 Oct;99(1):135-145 [FREE Full text] [doi: 10.1007/s11205-009-9572-0] [Medline: 20835391]
2. Wiggers K. Pew: Smartphone penetration ranges from 24% in India to 95% in South Korea. 2019. URL: https://venturebeat.com/2019/02/05/pew-south-korea-has-the-worlds-highest-smartphone-ownership-rate/ [accessed 2021-01-24]
3. Haug S, Castro RP, Kwon M, Filler A, Kowatsch T, Schaub MP. Smartphone use and smartphone addiction among young people in Switzerland. J Behav Addict 2015 Dec;4(4):299-307 [FREE Full text] [doi: 10.1556/2006.4.2015.037] [Medline: 26690625]
4. Seto E, Ross H, Tibbles A, Wong S, Ware P, Etchells E, et al. A Mobile Phone-Based Telemonitoring Program for Heart Failure Patients After an Incidence of Acute Decompensation (Medly-AID): Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2020 Jan 22;9(1):e15753 [FREE Full text] [doi: 10.2196/15753] [Medline: 32012116]
5. Debono B, Bousquet P, Sabatier P, Plas J, Lescure J, Hamel O. Postoperative monitoring with a mobile application after ambulatory lumbar discectomy: an effective tool for spine surgeons. Eur Spine J 2016 Nov;25(11):3536-3542. [doi: 10.1007/s00586-016-4680-4] [Medline: 27349754]
6. Skrepnik N, Spitzer A, Altman R, Hoekstra J, Stewart J, Toselli R. Assessing the Impact of a Novel Smartphone Application Compared With Standard Follow-Up on Mobility of Patients With Knee Osteoarthritis Following Treatment With Hylan G-F 20: A Randomized Controlled Trial. JMIR Mhealth Uhealth 2017 May 09;5(5):e64 [FREE Full text] [doi: 10.2196/mhealth.7179] [Medline: 28487266]
7. Armstrong KA, Coyte PC, Brown M, Beber B, Semple JL. Effect of Home Monitoring via Mobile App on the Number of In-Person Visits Following Ambulatory Surgery: A Randomized Clinical Trial. JAMA Surg 2017 Mar 22:622-627. [doi: 10.1001/jamasurg.2017.0111] [Medline: 28329223]
8. Gray RT, Sut MK, Badger SA, Harvey CF. Post-operative telephone review is cost-effective and acceptable to patients. Ulster Med J 2010 May;79(2):76-79 [FREE Full text] [Medline: 21116423]
9. Imam MA, Barke S, Stafford GH, Parkin D, Field RE. Loss to follow-up after total hip replacement: a source of bias in patient reported outcome measures and registry datasets? Hip Int 2014;24(5):465-472. [doi: 10.5301/hipint.5000141] [Medline: 24970318]

10. Sleat G, Kelly A. , Henry MB, Peter J. O. Predicting completion of follow-up in prospective orthopaedic trauma research. Orthopedics Trauma Association International 2019:e47. [doi: 10.1097/oi9.0000000000000047]
11. Whiting PS, Greenberg SE, Thakore RV, Alamanda VK, Ehrenfeld JM, Obremskey WT, et al. What factors influence follow-up in orthopedic trauma surgery? Arch Orthop Trauma Surg 2015 Mar;135(3):321-327. [doi: 10.1007/s00402-015-2151-8] [Medline: 25617213]
12. Connor MJ, Winkler M, Miah S. COVID-19 pandemic - is virtual urology clinic the answer to keeping the cancer pathway moving? BJU Int 2020 Jun;125(6):E3-E4. [doi: 10.1111/bju.15061] [Medline: 32232915]
13. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, et al. A standardized method for the assessment of shoulder function. J Shoulder Elbow Surg 1994 Nov;3(6):347-352. [doi: 10.1016/S1058-2746(09)80019-0] [Medline: 22958838]
14. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). Am J Ind Med 1996 Jun;29(6):602-608. [doi: 10.1002/(SICI)1097-0274(199606)29:6<602::AID-AJIM4>3.0.CO;2-L] [Medline: 8773720]
15. Paiva CE, Barroso EM, Carneseca EC, de Pádua Souza C, Dos Santos FT, Mendoza López RV, et al. A critical analysis of test-retest reliability in instrument validation studies of cancer patients under palliative care: a systematic review. BMC Med Res Methodol 2014 Jan 21;14:8 [FREE Full text] [doi: 10.1186/1471-2288-14-8] [Medline: 24447633]
16. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health 2009 Jun;12(4):419-429. [doi: 10.1111/j.1524-4733.2008.00470.x] [Medline: 19900250]
17. Chahal J, Whelan DB, Jaglal SB, Smith P, MacDonald PB, Levy BA, et al. The multiligament quality of life questionnaire: development and evaluation of test-retest reliability and validity in patients with multiligament knee injuries. Am J Sports Med 2014 Dec;42(12):2906-2916. [doi: 10.1177/0363546514552629] [Medline: 25288625]
18. Bramming IB, Kierkegaard S, Lund B, Jakobsen SS, Mechlenburg I. High relative reliability and responsiveness of the forgotten joint score-12 in patients with femoroacetabular impingement undergoing hip arthroscopic treatment. A prospective survey-based study. J Hip Preserv Surg 2019 Jul;6(2):149-156 [FREE Full text] [doi: 10.1093/jhps/hnz020] [Medline: 31660200]
19. Zelle BA, Bhandari M, Sanchez AI, Probst C, Pape H. Loss of follow-up in orthopaedic trauma: is 80% follow-up still acceptable? J Orthop Trauma 2013 Mar;27(3):177-181. [doi: 10.1097/BOT.0b013e31825cf367] [Medline: 23449099]
20. Casp AJ, Wells J, Holzgrefe R, Weiss D, Kahler D, Yarboro SR. Evaluation of Orthopedic Trauma Surgery Follow-up and Impact of a Routine Callback Program. Orthopedics 2017 Mar 01;40(2):e312-e316. [doi: 10.3928/01477447-20161229-01] [Medline: 28056157]
21. Cronin KH, Hockensmith L. The Invisible Patient: Defining the Orthopaedic Trauma Patient that Does Not Follow Up. 2018 Presented at: AAOS 2018 Annual Meeting; Unknown; New Orleans, LA.
22. Higgins J, Chang J, Hoit G, Chahal J, Dwyer T, Theodoropoulos J. Conventional Follow-up Versus Mobile Application Home Monitoring for Postoperative Anterior Cruciate Ligament Reconstruction Patients: A Randomized Controlled Trial. Arthroscopy 2020 Jul;36(7):1906-1916. [doi: 10.1016/j.arthro.2020.02.045] [Medline: 32268161]
23. Tang P, Hou Z, Wu X, Zhang C, Wang J, Xing X, Orthopedic Trauma Branch of Chinese Orthopedic Association; External FixationLimb Reconstruction Branch of Chinese Orthopedic Association; Trauma Expert Committee of Chinese Association of Orthopedic Surgeons. Expert consensus on management principles of orthopedic emergency in the epidemic of coronavirus disease 2019. Chin Med J (Engl) 2020 May 05;133(9):1096-1098 [FREE Full text] [doi: 10.1097/CM9.0000000000000810] [Medline: 32358326]
24. World Health Organization. Novel coronavirus (2019-nCoV) situation report - 1. 2020. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4 [accessed 2020-03-24]
25. (WHO) W. Coronavirus disease 2019 (COVID-19) situation report - 56. 2020. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200316-sitrep-56-covid-19.pdf?sfvrsn=9fda7db2_6 [accessed 2020-03-24]
26. (CDC) USCfDCaP. Coronavirus disease 2019 (COVID-19) - situation summary. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/summary.html [accessed 2020-03-17]
27. Vannabouathong C, Devji T, Ekhtiari S, Chang Y, Phillips SA, Zhu M, et al. Novel Coronavirus COVID-19: Current Evidence and Evolving Strategies. J Bone Joint Surg Am 2020 May 06;102(9):734-744 [FREE Full text] [doi: 10.2106/JBJS.20.00396] [Medline: 32379112]
28. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. Value Health 2008;11(2):322-333. [doi: 10.1111/j.1524-4733.2007.00231.x] [Medline: 18380645]
29. Yaffe M, Goyal N, Kokmeyer D, Merrell GA. The use of an iPad to collect patient-reported functional outcome measures in hand surgery. Hand (N Y) 2015 Sep;10(3):522-528 [FREE Full text] [doi: 10.1007/s11552-014-9731-x] [Medline: 26330789]

Abbreviations

ASES: American Shoulder and Elbow Society
DASH: Disabilities of the Arm, Shoulder, and Hand
ICC: intraclass correlation coefficient
PROMs: patient-reported outcome measures
VAS: visual analog scale

Edited by R Kukafka; submitted 05.06.20; peer-reviewed by H Söbke; comments to author 11.07.20; revised version received 16.08.20; accepted 24.01.21; published 01.03.21.

Please cite as:
Hong TH, Kim MK, Ryu DJ, Park JS, Bae GC, Jeon YS
The Reliability of Remote Patient-Reported Outcome Measures via Mobile Apps to Replace Outpatient Visits After Rotator Cuff Repair Surgery: Repetitive Test-Retest Comparison Study for 1-Year Follow-up
J Med Internet Res 2021;23(3):e20989
URL: https://www.jmir.org/2021/3/e20989
doi:10.2196/20989
PMID:33646133

Original Paper

Preferences for Artificial Intelligence Clinicians Before and During the COVID-19 Pandemic: Discrete Choice Experiment and Propensity Score Matching Study

Taoran Liu[1*], BSc; Winghei Tsang[2*], MBBS; Yifei Xie[2], MBBS; Kang Tian[3], BSc; Fengqiu Huang[1], MPH; Yanhui Chen[2], MBBS; Oiying Lau[2], MBBS; Guanrui Feng[1], MPH; Jianhao Du[1], MPH; Bojia Chu[4], BA, MSc; Tingyu Shi[3], BSc; Junjie Zhao[5], BSc; Yiming Cai[6], BEcon; Xueyan Hu[1], BEcon; Babatunde Akinwunmi[7,8], MD, MPH, MMSc; Jian Huang[9], MPH, PhD; Casper J P Zhang[10], MPH, PhD; Wai-Kit Ming[1], MD, PhD, MPH, MMSc

[1]Department of Public Health and Preventive Medicine, School of Medicine, Jinan University, Guangzhou, China

[2]International School, Jinan University, Guangzhou, China

[3]Faculty of Social Sciences, University of Southampton, Southampton, United Kingdom

[4]Department of Applied Mathmatics, The Hong Kong Polytechnic University, Hong Kong, Hong Kong

[5]College of Computer Science and Technology, Henan Polytechnic University, Henan, China

[6]School of Applied Mathematics, Beijing Normal University (Zhuhai), Zhuhai, China

[7]Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, MA, United States

[8]Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Harvard University, Boston, MA, United States

[9]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom

[10]School of Public Health, The University of Hong Kong, Hong Kong, Hong Kong

[*]these authors contributed equally

Corresponding Author:
Wai-Kit Ming, MD, PhD, MPH, MMSc
Department of Public Health and Preventive Medicine
School of Medicine
Jinan University
601 Huangpu W Ave
Tianhe District
Guangzhou, 510632
China
Phone: 86 85228852
Email: wkming@connect.hku.hk

Abstract

Background: Artificial intelligence (AI) methods can potentially be used to relieve the pressure that the COVID-19 pandemic has exerted on public health. In cases of medical resource shortages caused by the pandemic, changes in people's preferences for AI clinicians and traditional clinicians are worth exploring.

Objective: We aimed to quantify and compare people's preferences for AI clinicians and traditional clinicians before and during the COVID-19 pandemic, and to assess whether people's preferences were affected by the pressure of pandemic.

Methods: We used the propensity score matching method to match two different groups of respondents with similar demographic characteristics. Respondents were recruited in 2017 and 2020. A total of 2048 respondents (2017: n=1520; 2020: n=528) completed the questionnaire and were included in the analysis. Multinomial logit models and latent class models were used to assess people's preferences for different diagnosis methods.

Results: In total, 84.7% (1115/1317) of respondents in the 2017 group and 91.3% (482/528) of respondents in the 2020 group were confident that AI diagnosis methods would outperform human clinician diagnosis methods in the future. Both groups of matched respondents believed that the most important attribute of diagnosis was accuracy, and they preferred to receive combined diagnoses from both AI and human clinicians (2017: odds ratio [OR] 1.645, 95% CI 1.535-1.763; *P*<.001; 2020: OR 1.513, 95% CI 1.413-1.621; *P*<.001; reference: clinician diagnoses). The latent class model identified three classes with different attribute priorities. In class 1, preferences for combined diagnoses and accuracy remained constant in 2017 and 2020, and high accuracy

(eg, 100% accuracy in 2017: OR 1.357, 95% CI 1.164-1.581) was preferred. In class 2, the matched data from 2017 were similar to those from 2020; combined diagnoses from both AI and human clinicians (2017: OR 1.204, 95% CI 1.039-1.394; *P*=.011; 2020: OR 2.009, 95% CI 1.826-2.211; *P*<.001; reference: clinician diagnoses) and an outpatient waiting time of 20 minutes (2017: OR 1.349, 95% CI 1.065-1.708; *P*<.001; 2020: OR 1.488, 95% CI 1.287-1.721; *P*<.001; reference: 0 minutes) were consistently preferred. In class 3, the respondents in the 2017 and 2020 groups preferred different diagnosis methods; respondents in the 2017 group preferred clinician diagnoses, whereas respondents in the 2020 group preferred AI diagnoses. In the latent class, which was stratified according to sex, all male and female respondents in the 2017 and 2020 groups believed that accuracy was the most important attribute of diagnosis.

Conclusions: Individuals' preferences for receiving clinical diagnoses from AI and human clinicians were generally unaffected by the pandemic. Respondents believed that accuracy and expense were the most important attributes of diagnosis. These findings can be used to guide policies that are relevant to the development of AI-based health care.

(J Med Internet Res 2021;23(3):e26997) doi:10.2196/26997

KEYWORDS

propensity score matching; discrete latent traits; patients' preferences; artificial intelligence; COVID-19; preference; discrete choice; choice; traditional medicine; public health; resource; patient; diagnosis; accuracy

Introduction

Artificial intelligence (AI) technology, which is also called machine intelligence technology, has been used in various fields, such as automation, language, image understanding and analysis, and genetic algorithm research. AI technology can perform better than a human when it comes to performing particular tasks, and such technology has the potential to replace several traditional human occupations. This is the result of continuous advances in medicine, neuroscience, robotics, and statistics. In the medical and health care field [1], AI technology has many widespread applications, and the use of such technology has resulted in a wide range of opportunities for the future. For instance, machine learning technology has been used to analyze medical big data and electronic health records, conduct computer vision research, facilitate natural language processing, and develop intelligent robots [2]. In addition, AI technology has helped address the masses' demands for increasing the number of clinician services [3].

As of November 13, 2020, the novel COVID-19 disease has spread in over 217 countries [4] and territories across the world. The disease has also resulted in tremendous threats and challenges to public health security systems worldwide. The COVID-19 outbreak has pushed the medical systems and resources of numerous countries to the brink of collapse. Diagnostic AI technology, which includes diagnostic machine learning technology, has started to play a role in relieving the burden that the pandemic has placed on the public health system and easing the shortage of medical resources. At the start of the COVID-19 outbreak, the medical AI team of the Alibaba Academy for Discovery, Adventure, Momentum, and Outlook rapidly developed a set of AI diagnostic technologies that could interpret the computed tomography images of patients with suspected COVID-19 (ie, new COVID-19 cases) within 20 seconds, with an accuracy of 96% [5]. In the fight against the epidemic [6], digital technologies such as cloud computing, artificial intelligence, and blockchain technologies have played a vital role.

The combination of AI technology and human clinician–operated convolutional neural networks [7] has greatly improved the efficiency and accuracy of diagnosis methods and substantially reduced diagnosis times and outpatient queuing times. In 2014, app developers from around the world made a total of US $663.8 million by selling AI health care apps, and their revenue is expected to reach US $666.2 million in 2021 [8]. However, there are various uncertainties with regard to preferences for different diagnostic methods among patients (ie, men and women) from high-income areas and low-income areas in China. Furthermore, there have been no studies that assess patients' preferences for AI clinicians and human clinicians before and during the COVID-19 pandemic period, and analyze the aspects of patients' decision-making behaviors during different periods of time.

This study aimed to compare people's preferences for AI diagnoses and traditional diagnoses (ie, human clinicians' diagnoses) before and during the COVID-19 pandemic. We assessed two groups of respondents with similar demographic characteristics. We recruited one group in 2017 and the other group in 2020 to learn whether people's preferences for AI and traditional human clinicians were affected by the pressure of the COVID-19 pandemic. We performed propensity score matching (PSM) to match the two groups. We also conducted a discrete choice experiment (DCE) to quantify and measure peoples' preferences for different diagnosis methods and identify factors that disrupted and impacted peoples' decision-making behaviors.

Methods

Overview

We designed a web-based questionnaire to collect participants' demographic information and investigate patients' preferences for different diagnosis strategies (Multimedia Appendix 1). In brief, the questionnaire included 7 similar hypothetical scenarios. Respondents were asked to choose a preferred diagnosis strategy for each scenario.

We used the PSM method to match two different groups of respondents (ie, the 2017 group and the 2020 group) with similar demographic characteristics. In addition, we used multinomial logit (MNL) models [9,10] and latent class models (LCMs) [11]

to evaluate and investigate respondents' preferences for different diagnosis strategies. We also compared the preferences of the matched respondents from the 2017 group to those of the 2020 group to identify heterogeneity or homogeneity in preferences for diagnosis attributes.

Selection of Attributes and Levels

Individuals could choose different levels of health care services for each diagnosis attribute. Patients from the outpatient queues of The First Affiliated Hospital of Jinan University (Guangzhou Overseas Chinese Hospital) and The First Affiliated Hospital of Sun Yat-sen University were randomly selected for this study. Each patient was prompted to hypothesize which diagnosis methods or attributes had a large impact on their decision (ie, the methods/attributes that were of prominent importance to each participant).

After assessing patients' hypotheses and related literature [12-14], we included the following six diagnosis attributes and their respective levels in our questionnaire experiment: (1) diagnostic method (levels: clinician diagnosis, AI and clinician diagnosis, and AI diagnosis); (2) outpatient waiting time before the start of the diagnosis process (levels: 0, 20, 40, 60, 80, and 100 minutes); (3) diagnosis time (levels: 0, 15, and 30 minutes); (4) accuracy (ie, the rate of correct diagnosis; levels: 60%, 70%, 80%, 90%, and 100%); (5) follow-up after diagnosis (ie, whether a doctor can conduct follow-ups at any time; levels: yes or no); and (6) diagnostic expenses (levels: ¥0, ¥50, ¥100, ¥150, ¥200, and ¥250; a currency exchange rate of ¥1=US $0.16 is applicable). Attributes and their respective levels are presented in Textbox 1.

Textbox 1. Diagnosis attributes and their respective levels in this discrete choice experiment.

Diagnostic method

- Description: the diagnosis method that patients prefer
- Levels: clinician diagnosis, artificial intelligence and clinician diagnosis, and artificial intelligence diagnosis

Outpatient waiting time

- Description: the amount of time that patients wait in a queue before the diagnosis process
- Levels: 0 minutes, 20 minutes, 40 minutes, 60 minutes, 80 minutes, and 100 minutes

Diagnosis time

- Description: the amount of time before a patient obtains a diagnosis
- Levels: 0 minutes, 15 minutes, and 30 minutes

Diagnostic accuracy

- Description: the rate of correct diagnosis
- Levels: 60%, 70%, 80%, 90%, and 100%

Follow-up after diagnosis

- Description: case tracking and follow-ups after diagnosis
- Levels: Yes and no

Diagnostic expenses

- Description: the cost of diagnosis
- Levels: ¥0, ¥50, ¥100, ¥150, ¥200, and ¥250 (a currency exchange rate of ¥1=US $0.16 is applicable)

DCE Instrument Design and Questionnaire

With regard to the design our DCE instrument, we used the fractional factorial design method [15,16] to identify the optimal number of treatment scenarios. This process was conducted with Lighthouse Studio version 9.8.1 (Sawtooth Software). In practice, it is not always feasible for respondents to choose among all of the possible combinations of attributes and levels (ie, full factorial design). The full factorial design of the DCE instrument had 3240 different combinations (ie, $3 \times 6 \times 3 \times 5 \times 2 \times 6 = 3264$), which is an unreasonable number of options to present to respondents. Thus, the fractional factorial method was essential in designing the DCE instrument. This method is based on the following two principles [15-17]: (1) orthogonality, which, in terms of the DCE, means that each attribute level should have little to no correlation with other attribute levels; and (2) balance, which means that each attribute should appear an equal number of times. After considering these principles, we provided 6 random questions and 1 fixed question to each respondent in the DCE.

The DCE questionnaire contained 2 parts. The first part required the respondents to fill in their demographic information, such as age (ie, 18-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, 61-65, 66-70, 71-75, 76-80, and 81-85 years), sex (ie, male or female), and educational level (ie, primary school student, primary school graduate, middle school student, middle

school graduate, high school student, high school graduate, undergraduate, bachelor's degree, graduate student, master's degree, postgraduate student, and doctorate degree). The second part required the respondents to consider seven different scenarios. For each scenario, respondents were to imagine that they were in an outpatient queue waiting for a diagnosis. They were then asked to choose a preferred diagnosis strategy. At the end of the questionnaire, respondents were required to estimate the number of years (ie, 5 years, 10 years, 15 years, 20 years, 30 years, 40 years, or never) it would take for AI clinicians to surpass human clinicians. The scenarios and the options for the different types of clinicians are presented in Multimedia Appendix 2.

Data Collection

In October 2017 and August 2020, we sent our website link to people of different age groups by using various social media platforms, such as WeChat (Tencent Inc) and QQ (Tencent Inc). People could use the link to access the DCE questionnaire, which was the same for each participant. To increase the response rate, we provided incentives (ie, a lottery for a Fitbit watch and cash prizes) for completing the questionnaire.

At the beginning of the questionnaire, we provided a brief background on the applications of AI in medicine. This included information on the potential advantages and disadvantages of AI clinicians and traditional clinicians, and the purpose of our DCE. The questionnaire only took 5-10 minutes to complete. Respondents had to click the "Agree to take the survey" button to start filling out the questionnaire. Once respondents clicked the "Agree to take the survey" button, they were notified that they willingly chose to participate in this study. Respondents were also notified that their privacy was protected by the law.

PSM

PSM is a regression method for identifying treatment group and control group patients with similar basic characteristics. This method is prevalently used in the study of impact factors and causal effects, such as those in medical treatments, policy decisions, or case studies. PSM involves the following five steps [18]: (1) estimating propensity scores; (2) choosing a matching algorithm; (3) checking for overlap/common support; (4) estimating the quality and effects of the matching results; and (5) conducting a sensitivity analysis. The mathematical theory for PSM is primarily based on the Roy-Robin model [19-21]. Our objective was to perform a PSM analysis in which participants who were recruited in 2017 were treated as the treatment group, and participants who were recruited in 2020 were treated as the control group. Participants' PSM data are provided in Multimedia Appendix 3 [18]. We matched the respondents in each group according to their demographic characteristics, such as age, sex, and educational level. All demographic information was coded as dummy variables; for instance, male respondents were coded as "1," and female respondents were coded as "0."

Matching Algorithm

Although there are various matching algorithms [18], we used the nearest neighbor [22] algorithm because it was appropriate for identifying individuals in one group that best matched the individuals in another group. Another merit of the nearest neighbor algorithm is that it can differentiate between individuals in the control group and individuals in the treatment group, which guarantees that all treated individuals are successfully matched. Therefore, the nearest neighbor algorithm provides the most information on treatment groups and control groups. Additionally, we conducted a 1:1 matching analysis, which effectively reduces confounding bias [23] and improves research efficiency and credibility.

Statistical Analysis

MNL Model

There are various analysis models that can be used to conduct DCE-related statistical analyses, such as random effects binary probit and logit models, MNL models, and mixed logit models [16,24]. The theoretical model for a DCE is based on the random utility model (Multimedia Appendix 4) [16]. We assumed that respondents' choices would maximize the utility of each question in the DCE questionnaire. The overall utility of decision makers is based on fixed utility and random utility, which are unobservable. We assessed respondents' preferences by analyzing their comments. This allowed us to identify random utilities that could not be identified by analyzing a question.

We used the MNL model to analyze people's preferences for different attribute levels. Our independent variable only accounted for attributes that were related to health care plans; it did not account for any information that was related to participants. The MNL model was used to analyze respondents' health care plans, which were chosen based on the relative importance of the plans' attributes and the "none" option. The coded value of each participants' chosen health care plan was calculated based on participants' coded responses to questions about queuing times, diagnosis times, and diagnostic costs. We used a maximum likelihood approach to analyze MNL model data.

The results from the MNL model were determined by the options for health care plans, as the data for this attribute were grouped before analysis. In the MNL model, "effect" is synonymous with "utility." Therefore, positive MNL model coefficients indicated that individuals preferred one level of service over other levels for the same attribute. The MNL model in this study was based on a similar logistic regression model. The MNL model–based observations correlated with those in blocks that corresponded with the same individual. Instead of having 1 level line per individual like in the classical logit model, the MNL model had 1 level line per attribute level of interest (ie, for each individual). For example, in this study, we analyzed three types of diagnoses (ie, clinician diagnoses, AI and clinician diagnoses, and AI diagnoses), and each type had its own characteristics. However, an individual could only choose 1 of the 3 types of diagnoses. As per the characteristics of the MNL model, all three options were presented to each respondent, and all respondents could choose their preferred option. We reported the odds ratios (ORs) of respondents' preferences for different attribute levels.

LCM

We used an LCM [11] to create different classes for individuals with similar preferences. The purpose of the LCM was to identify correlations among explicit variables, create the fewest number of classes, and achieve local independence. An LCM initially assumes that the null model is the hypothesized model and that local independence exists among explicit variables. Afterward, the LCM increases the number of latent categories in the null model and uses a maximum likelihood approach to create various models, which are based on parameters' limitations. The LCM then tests the hypothesized model and observed data, compares the hypothesized model to the other models, and identifies the most appropriate model. Although there are different types of model information evaluation criteria, Akaike information criteria [25] and Bayesian information criteria [26] are the most prevalently used criteria for selecting LCMs. After the model was created, observed data were classified into the appropriate latent classes.

Willingness to Pay

Willingness to pay (WTP) is an efficient metric for measuring how much an individual is willing to sacrifice (ie, economic sacrifices) to choose one diagnosis attribute level over another (ie, the reference attribute level). We analyzed participants' WTP to identify homogeneity and heterogeneity in participants' preferences.

Software

Propensity score matching was conducted with Stata 16 (StataCorp LLC), and the MNL model and LCMs were created with Lighthouse Studio version 9.8.1 (Sawtooth Software).

Results

Data Collection

Of the 1520 individuals who visited our DCE website in 2017, 1317 (86.6%) completed the questionnaire and were included in the analysis. Of these 1317 respondents, 1317 (100%) were aged 18-85 years, 731 (55.5%) were female, and 1115 (84.7%) believed that AI clinicians would surpass or replace human clinicians.

Of the 874 individuals who visited our new DCE website in 2020, 528 (60.4%) completed the questionnaire. Of these 528 participants, 272 (51.5%) were female and 482 (91.3%) were confident that AI diagnoses were better than traditional diagnoses.

General PSM and MNL Model Results

Of the 1317 respondents who were recruited in 2017, 528 (40.1%) were matched (ie, via PSM) to the 528 respondents who were recruited in 2020. The PSM procedure is presented in Figure 1, and the demographic characteristics of respondents before and after PSM are presented in Table 1. The general MNL model results for the 2017 and 2020 groups are presented in Table 2, which shows estimated average preference weights (ie, effect weights), *P* values, ORs, and 95% confidence intervals. Generally, individuals in the 2017 and 2020 groups believed that accuracy was the most important diagnosis attribute (Figure 2). The weighted importance value of accuracy was 38.53% in the 2017 group and 40.55% in the 2020 group. Respondents believed that diagnosis time was the least important attribute (weighted importance in 2017: 2.69%; weighted importance in 2020: 1.16%). Additionally, individuals in the 2017 and 2020 groups preferred to receive combined diagnoses from both AI and human clinicians over AI-only diagnoses or human clinician–only diagnoses (2017: OR 1.645, 95% CI 1.535-1.763; 2020: OR 1.513, 95% CI 1.413-1.621; reference: clinician diagnosis; Table 2). In addition, the ORs for the levels of diagnosis accuracy increased as the accuracy increased, which indicated that people will always prefer diagnosis methods with high accuracy. For instance, in the 2017 group, 100% accuracy had an OR of 5.043 (95% CI 4.534-5.609). In the 2020 group, 100% accuracy had an OR of 5.263 (95% CI 4.734, 5.852). The preferences of the matched respondents in the 2017 group were very similar to those of the respondents in the 2020 group.

Figure 1. Propensity score matching procedure.

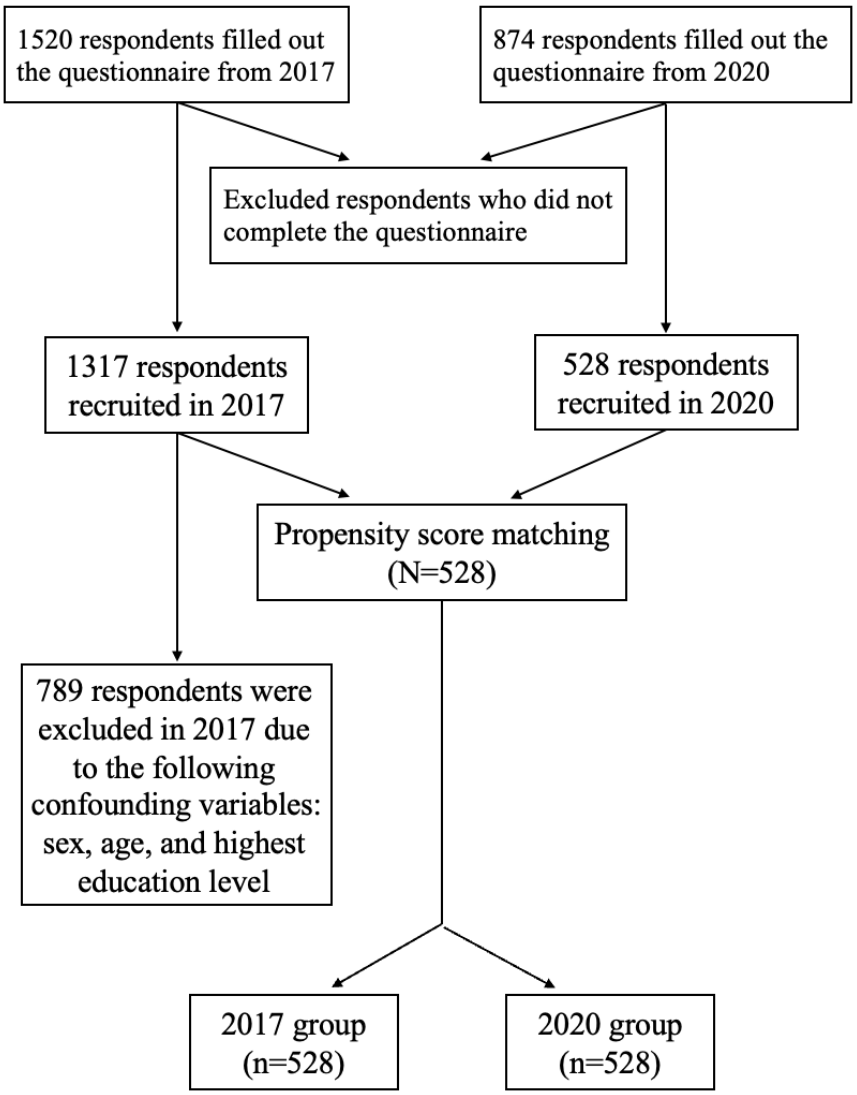


Table 1. Demographic characteristics of nonmatched and propensity score–matched respondents.

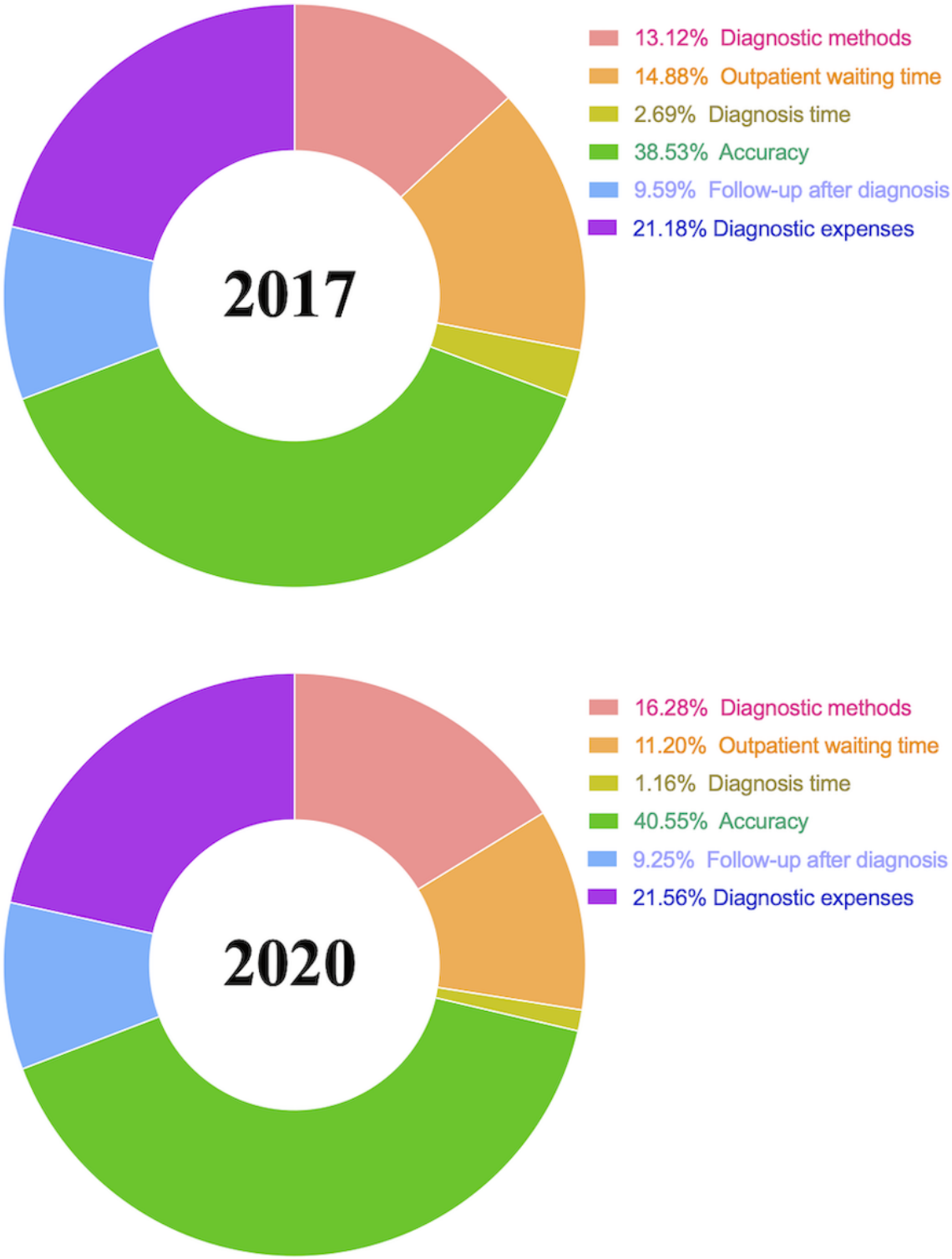
Baseline matching characteristics	Nonmatched respondents			Propensity score–matched respondents		
	2017 group (n=1317), n (%)	2020 group (n=528), n (%)	*P* value	2017 group (n=528), n (%)	2020 group (n=528), n (%)	*P* value
Sex			<.001			.97
Male	586 (44.5)	256 (48.48)		250 (47.35)	256 (48.48)	
Female	731 (55.5)	272 (51.52)		278 (52.65)	272 (51.52)	
Age (years)			<.001			.69
<35	1106 (83.98)	348 (65.91)		379 (71.78)	348 (65.91)	
≥35	211 (16.02)	180 (34.09)		149 (28.22)	180 (34.09)	
Highest education level			<.001			.13
Primary school graduate to undergraduate	1033 (78.44)	336 (63.64)		385 (72.92)	336 (63.64)	
Bachelor's degree to doctorate degree	284 (21.56)	192 (36.36)		143 (27.08)	192 (36.36)	

Table 2. General results of the multinomial logit model. Data on propensity score–matched respondents' preferences for diagnosis attributes in 2017 and 2020 are reported (N=528).

Attributes and levels	2017 group			2020 group		
	Effect coefficient	*P* value	Odds ratio (95% CI)	Effect coefficient	*P* value	Odds ratio (95% CI)
Diagnosis method						
Clinician	−0.15	<.001	Reference	−0.05	.12	Reference
Artificial intelligence and clinician	0.35	<.001	1.64 (1.535-1.763)	0.36	<.001	1.51 (1.413-1.621)
Artificial intelligence	−0.20	<.001	0.95 (0.885-1.016)	−0.31	<.001	0.78 (0.725-0.833)
Outpatient waiting time (minutes)						
0	0.31	<.001	Reference	0.15	.01	Reference
20	0.12	.03	0.82 (0.741-0.914)	0.26	<.001	1.12 (1.013-1.245)
40	−0.03	.57	0.71 (0.639-0.789)	−0.02	.72	0.85 (0.762-0.942)
60	−0.08	.12	0.67 (0.606-0.748)	−0.20	<.001	0.71 (0.640-0.788)
80	−0.31	<.001	0.54 (0.482-0.595)	−0.20	<.001	0.71 (0.640-0.789)
Diagnosis time (minutes)						
0	0.05	.19	Reference	−0.02	.57	Reference
15	−0.07	.06	0.89 (0.834-0.957)	−0.01	.83	1.01 (0.946-1.084)
30	0.02	.53	0.98 (0.912-1.046)	0.03	.43	1.05 (0.980-1.122)
Diagnosis accuracy (% accuracy)						
60	−0.83	<.001	Reference	−0.83	<.001	Reference
70	−0.35	<.001	1.62 (1.458-1.802)	−0.41	<.001	1.52 (1.365-1.684)
80	0.07	.16	2.47 (2.235-2.737)	−0.02	.72	2.25 (2.033-2.487)
90	0.32	<.001	3.18 (2.867-3.526)	0.43	<.001	3.51 (3.169-3.891)
100	0.79	<.001	5.04 (4.534-5.609)	0.83	<.001	5.26 (4.734-5.852)
Follow-up after diagnosis						
Yes	0.20	<.001	Reference	0.19	<.001	Reference
No	−0.20	<.001	0.67 (0.620-0.698)	−0.19	<.001	0.69 (0.656-0.715)
Diagnosis expenses (¥[a])						
0	0.42	<.001	Reference	0.36	<.001	Reference
50	0.28	<.001	0.87 (0.769-0.976)	0.23	<.001	0.88 (0.782-0.989)
100	−0.01	.82	0.65 (0.576-0.730)	0.18	<.001	0.83 (0.738-0.935)
150	0.03	.66	0.67 (0.599-0.760)	−0.06	.30	0.65 (0.580-0.736)
200	−0.24	<.001	0.52 (0.459-0.585)	−0.19	<.001	0.58 (0.510-0.648)
250	−0.47	<.001	0.41 (0.363-0.465)	−0.52	<.001	0.41 (0.366-0.468)

[a]A currency exchange rate of ¥1=US $0.16 is applicable.

Figure 2. General estimated weighted importance of diagnosis attributes in 2017 and 2020.



Overall WTP

In 2017, respondents were willing to pay ¥13.99 to receive combined diagnoses from AI and human clinicians. Additionally, people were not willing to pay for longer outpatient waiting times, but they were willing to pay for higher diagnosis accuracy (ie, ¥1.60 per 1% increase in accuracy). In 2020, respondents were willing to pay ¥0.79 to receive combined diagnoses from AI and human clinicians instead of clinician-only diagnoses. Compared to respondents' WTP for certain diagnosis methods in 2017, respondents' WTP in 2020 was lower. Furthermore, similar to the 2017 group, respondents in the 2020 group were also not willing to pay for longer outpatient waiting times. However, they were willing to pay for higher diagnosis accuracy.

LCM Results

After comparing the Akaike information criteria, Bayesian information criteria, and Akaike/Bayesian information criteria of the various potential classes, we chose three classes that were the most appropriate for the matched respondents in the 2017 and 2020 groups. The proportions of matched respondents from the 2017 group in each of the three classes were 43.2% (class 1: 228/528), 42.2% (class 2: 223/528) and 14.6% (class 3: 77/528). The proportions of matched respondents from the 2020 group in each of the three classes were 44.8% (class 1: 237/528), 48.2% (class 2: 254/528) and 7% (class 3: 37/528).

With regard to class 1 (n=228), Figure 3 shows that matched respondents in the 2017 group believed that diagnosis method was the most important attribute (weighted importance: 32.95%), followed by diagnosis expenses (weighted importance: 18.14%). In class 2, matched respondents from the 2017 group believed that diagnosis accuracy (weighted importance: 49.92%) and diagnosis expenses (weighted importance: 19.84%) were the most important attributes. In class 3, matched respondents from the 2017 group believed that diagnosis accuracy (weighted importance: 25.66%) and diagnosis expenses (weighted importance: 23.21%) were the most important attributes. In class 1, the respondents from the 2020 group believed that diagnosis expenses (weighted importance: 29.99%) and diagnosis method (weighted importance: 28.99%) were the most important attributes. In class 2, the respondents from the 2020 group believed that diagnosis accuracy (weighted importance: 52.34%) was the most important attribute, followed by diagnosis expenses (weighted importance: 14.44%). In class 3, the respondents from the 2020 group believed that diagnosis expense (weighted importance: 36.21%) was the most important attribute, followed by diagnosis accuracy (weighted importance: 32.84%). It was obvious that the three factors that respondents believed were the most important were diagnosis accuracy, diagnosis expenses, and diagnosis methods. In some classes, respondents believed that diagnosis method was the most important attribute. However, respondents typically believed that diagnosis accuracy was the most important attribute and diagnosis expense was the second most important attribute.

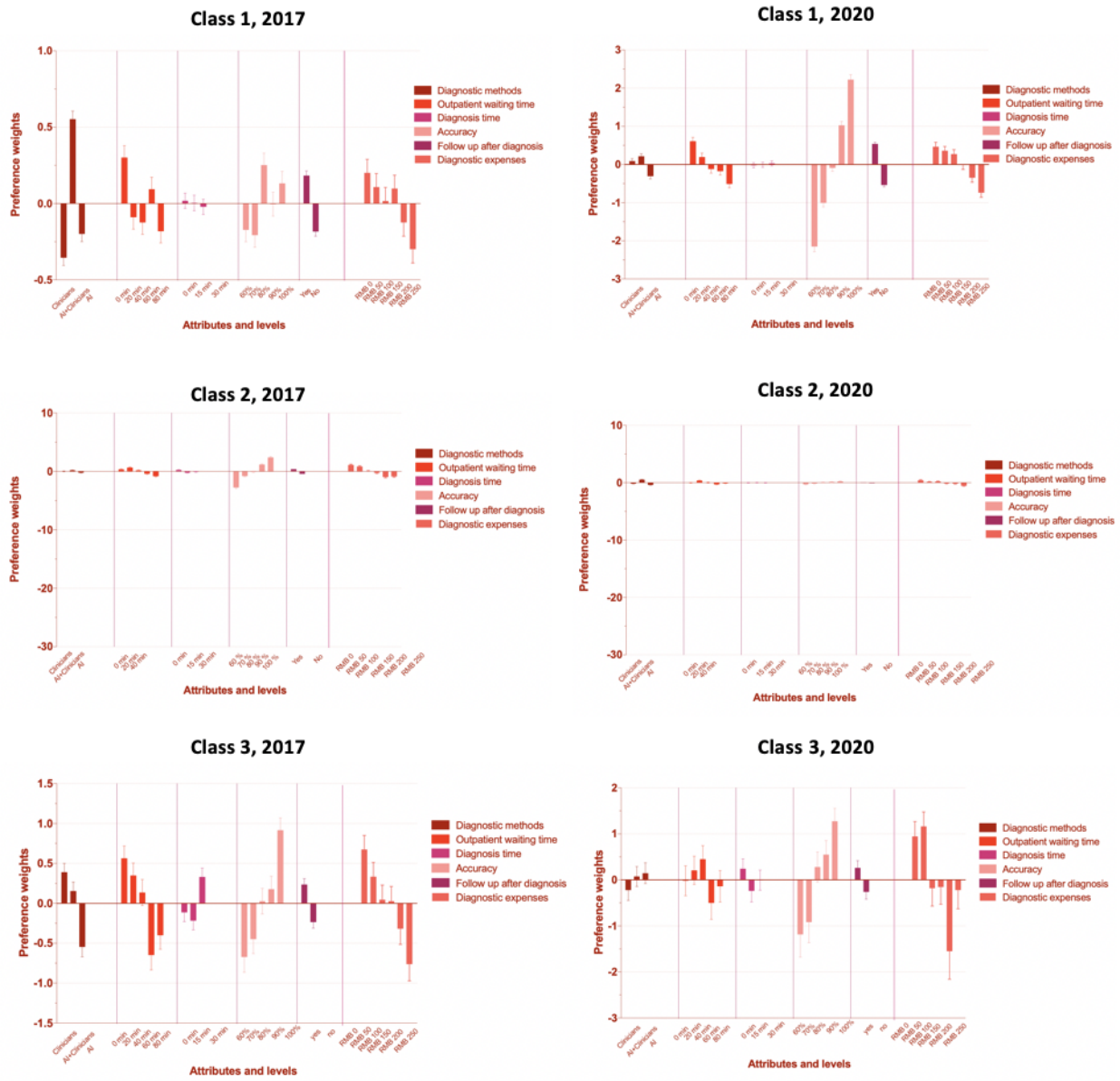
Figure 3. Weighted importance of diagnosis attributes in 2017 and 2020, as determined by the latent class model.

According to our ORs for classes 1 and 2, the respondents in the 2017 group (Table S1 in Multimedia Appendix 5) preferred the combined diagnosis method (class 1: OR 2.479, 95% CI 0.997-2.743; class 2: OR 1.204, 95% CI 1.039-1.394) over the other two methods. This was not true for respondents in class 3. Respondents in classes 1 and 3 preferred an outpatient waiting time of 0 minutes, and respondents in classes 1 and 2 preferred a diagnosis time of 0 minutes. Respondents across all classes preferred a diagnosis cost of ¥0. Furthermore, respondents in the 2017 group (ie, those in all classes) preferred high diagnosis accuracy (eg, 100% accuracy in class 3: OR 4.899, 95% CI 3.631-6.611). Respondents in all classes believed that follow-ups after diagnosis were important.

In classes 1 and 2, the respondents from the 2020 group (Table S2 in Multimedia Appendix 5) preferred the combined diagnosis method (class 1: OR 1.135, 95% CI 0.997-1.293; class 2: OR 2.009, 95% CI 1.826-2.211). This was not true for class 3. Respondents in class 2 preferred an outpatient waiting time of

20 minutes (OR 1.488, 95% CI 1.287-1.721). Additionally, similar to the 2017 group, the respondents in the 2020 group (ie, those in all classes) preferred high accuracy. Follow-ups after diagnosis were important to the respondents in the 2020 group (ie, those in all classes). The strength of respondents' preferences is visually presented in Figure 4; preference strength was quantified by calculating the preference weight (ie, coefficient) of each attribute's level.

Figure 4. Preference weights stratified by year (ie, 2017 and 2020) and class (ie, classes 1, 2, and 3), as determined by the latent class model.



We found that respondents' WTP was highly consistent with the corresponding ORs of each attribute. In classes 1 and 2, the respondents from the 2017 group (Table 3) were willing to pay for the combined diagnosis method. This was not true for class 3. Additionally, in class 3, the respondents from the 2017 group were the only respondents who were willing to pay for longer diagnosis times. The respondents from the 2017 group (ie, those in all classes) were willing to pay for higher diagnosis accuracy and follow-ups after diagnosis.

In classes 1 and 2, the respondents from the 2020 group (Table 4) were willing to pay for the combined diagnosis method. This was not true for class 3, in which respondents were willing to pay more for the AI diagnosis method. The respondents from the 2020 group (ie, those in all classes) were willing to pay for shorter outpatient waiting times, higher diagnosis accuracy, and follow-ups after diagnosis.

Table 3. Respondents' WTP[a] in 2017.[b]

Attribute	Overall WTP (N=528), ¥ (US $)	WTP in class 1 (n=228), ¥ (US $)	WTP in class 2 (n=223), ¥ (US $)	WTP in class 3 (n=77), ¥ (US $)
Diagnosis method				
Artificial intelligence and clinician	−13.99 (−2.24)	−3.03 (−0.48)	−0.22 (−0.04)	0.31 (0.05)
Artificial intelligence	1.50 (0.24)	−0.52 (−0.08)	0.25 (0.04)	1.22 (0.20)
Outpatient waiting time	8.92 (1.43)	0.62 (0.10)	0.96 (0.15)	0.53 (0.09)
Diagnosis time	−0.57 (−0.09)	0.07 (0.01)	0.07 (0.01)	−0.44 (−0.07)
Diagnosis accuracy	−1.14 (−0.18)	−0.44 (−0.07)	−2.85 (−0.46)	−1.20 (−0.19)
Follow-up after diagnosis	11.32 (1.81)	1.22 (0.20)	0.95 (0.15)	0.62 (0.10)
Diagnosis expenses	Reference	Reference	Reference	Reference

[a]WTP: willingness to pay.

[b]Negative currency values refer to the amount that respondents were willing to pay for another level.

Table 4. Respondents' WTP[a] in 2020.[b]

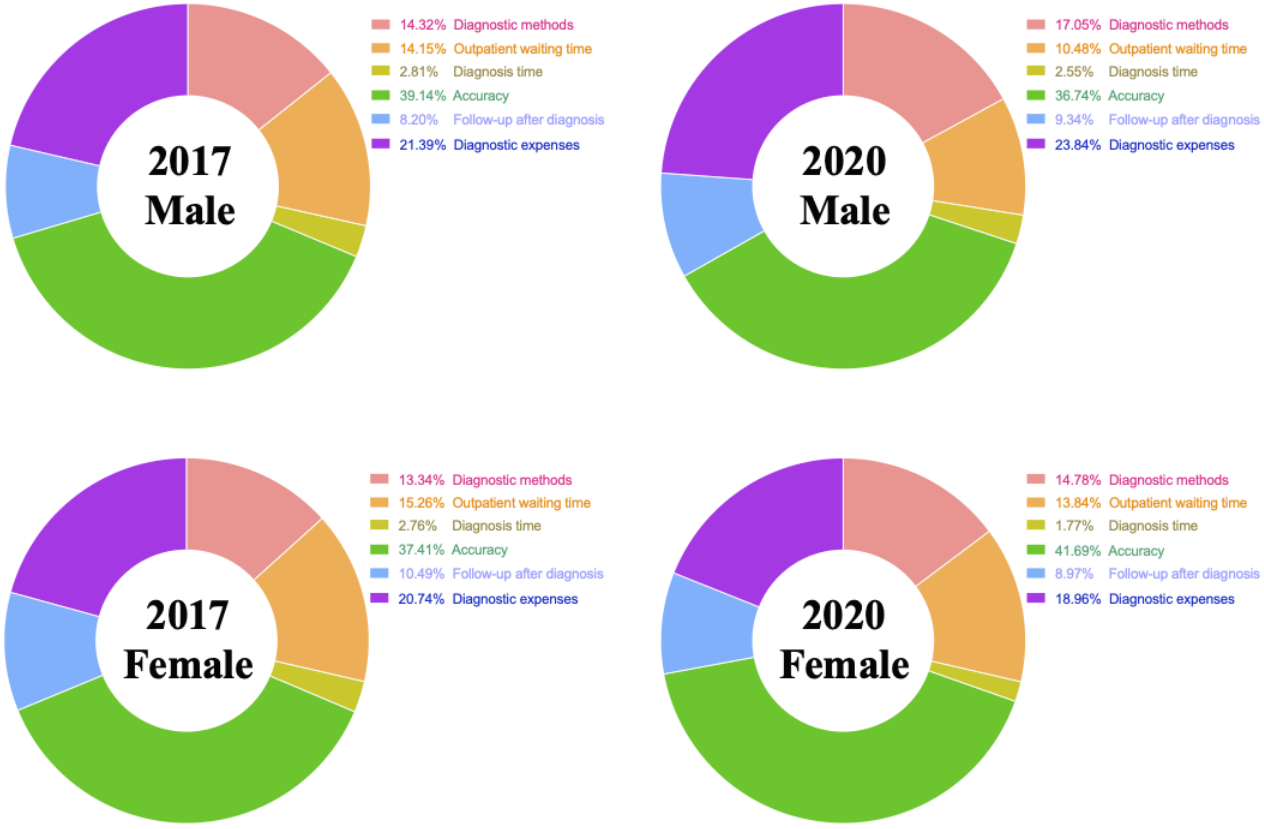
Attribute	Overall WTP (N=528), ¥ (US $)	WTP in class 1 (n=237), ¥ (US $)	WTP in class 2 (n=254), ¥ (US $)	WTP in class 3 (n=37), ¥ (US $)
Diagnosis method				
Artificial intelligence and clinician	−0.79 (−0.13)	−0.17 (−0.03)	−1.33 (−0.21)	−1.31 (−0.21)
Artificial intelligence	0.48 (0.07)	0.54 (0.09)	0.42 (0.07)	−1.62 (−0.26)
Outpatient waiting time	0.38 (0.06)	0.70 (0.11)	0.19 (0.03)	0.61 (0.10)
Diagnosis time	−0.05 (−0.01)	−0.04 (−0.01)	0.004 (0.001)	0.06 (0.01)
Diagnosis accuracy	−1.60 (−0.26)	−3 (−0.48)	−0.44 (−0.07)	−5.65 (−0.90)
Follow-up after diagnosis	0.73 (0.12)	1.46 (0.23)	0.25 (0.04)	2.31 (0.37)
Diagnosis expenses	Reference	Reference	Reference	Reference

[a]WTP: willingness to pay.

[b]Negative currency values refer to the amount that respondents were willing to pay for another level.

According to the LCM, which stratified data according to sex, male respondents in the 2017 group (Figure 5) believed that the most important attribute was diagnosis accuracy (weighted importance: 39.14%), followed by diagnosis expenses (weighted importance: 21.39%). Female respondents in the 2017 group also thought that diagnosis accuracy (weighted importance: 37.41%) and diagnosis expenses (weighted importance: 20.74) were the most important attributes. Male respondents in the 2020 group thought that diagnosis accuracy (weighted importance: 36.74%) was the most important attribute, followed by diagnosis expenses (weighted importance: 23.84%). Additionally, female respondents in the 2020 group believed that diagnosis accuracy (weighted importance: 41.69%) was the most important attribute, followed by diagnosis expenses (18.96%). The LCM for male and female respondents in the 2017 and 2020 groups showed that there was no obvious heterogeneity among these respondents' preferences.

Figure 5. Weighted importance of diagnosis attributes in 2017 and 2020, as determined by the latent class model, which stratified data according to sex (ie, male and female respondents).



Discussion

Principal Results

In this study, we collected information on people's preferences for AI-based diagnosis by analyzing two different groups of individuals who were recruited in 2017 and 2020 (ie, before and during the COVID-19 pandemic). We used the PSM method to match two groups of respondents with similar demographic characteristics (ie, age, sex, and educational level). After comparing the demographically similar respondents in the 2017 and 2020 groups, we did not find any substantial differences in respondents' preferences. Diagnosis accuracy and diagnosis expenses were the most important factors that influenced respondents' preferences.

The success of a DCE questionnaire always depends on the response rate. In other words, people who actively click the website link and complete the questionnaire are essential for expanding sample sizes and the scope of a study. By using the PSM method, we were able to easily assess whether people's preferences during normal times changed during unusual times (ie, the COVID-19 pandemic).

In this study, we used two different models—the MNL model and the LCM. Both models have various advantages and drawbacks with regard to quantifying respondents' preferences. According to the general PSM logit model, respondents in both groups consistently believed that accuracy was the most important diagnosis attribute, regardless of their preferences for diagnosis methods. Moreover, diagnosis expense was an important factor that influenced respondents' decisions in both 2017 and 2020. Respondents believed that this attribute was the second most important attribute. The limited accessibility and availability of medical resources are big problems in China, especially in several rural areas of China. These problems are the result of insufficient medical insurance distribution [27,28] and the country's low per capita income.

We found that people's preferences for different diagnoses were largely similar. This indicates that people's decisions and their preferences for different diagnoses are not considerably affected by pandemic-related factors. However, according to our LCM, there was slight heterogeneity in the preferences of different groups of respondents (eg, male and female respondents). This heterogeneity was not observed in the logit model. Although the weighted importance of accuracy remained consistent across all classes, it might not be the most important factor that affects people's decisions. In class 1, the respondents from the 2017 and 2020 groups believed that diagnosis expense was the most important factor that affected their decisions, followed by diagnosis method. Based on the LCM results, male respondents in the 2017 and 2020 groups believed that diagnosis accuracy was the most important attribute to consider when choosing a diagnosis strategy.

With regard to attribute levels, we found that respondents typically preferred to receive a combined diagnosis from both AI and human clinicians over a diagnosis from a single source (ie, AI diagnoses or human clinician diagnoses). This is understandable, since respondents typically believed that diagnosis accuracy could be improved by combining different

modes of diagnosis. Additionally, it should be noted that several respondents preferred longer diagnosis and outpatient queuing times. Although no studies have reported that diagnosis time and outpatient time correlate with diagnosis accuracy, it is possible that some patients prefer waiting for a doctor over receiving a quicker diagnosis, as they may believe that waiting results in more accurate diagnoses. The low accessibility and high price of AI services are important issues, especially in rural or low-income areas. Therefore, before pricing an AI technology–based service, it is advisable to survey residents and analyze their disposable income. With regard to residents in rural areas, governments should consider adding AI diagnoses to health insurance plans or related subsidy projects. Another AI diagnosis factor that should be considered is accuracy, since companies should only promote and advertise products/services with a high accuracy. When an AI technology–based service enters the market, relevant users should consider combining AI technology with human wisdom during the early stage of market penetration. Therefore, in the future, AI diagnosis technology developers should focus on improving diagnosis accuracy and reducing the cost of diagnoses to make such technology accessible to a wide range of patients.

Limitations

Our study has several shortcomings and limitations, especially with regard to our data collection process. It was clear that our small sample size limited the power of our analyses. Additionally, our sample might not be representative of the entire Chinese population. Furthermore, the deployment/distribution of AI technology–based medical services is limited, especially in rural areas [29] and areas that consist of uneducated residents. Thus, there are still many obstacles to overcome before AI technology becomes popular; many developments are still needed to popularize conceptual projects.

Conclusion

Our study shows that respondents' preferences for AI clinicians in 2017 did not substantially differ from those in 2020. Therefore, people's preferences for AI diagnoses and clinical diagnoses were unaffected by the COVID-19 pandemic. However, preferences for high diagnostic accuracy and low diagnosis expenses were evident, regardless of people's preferences for diagnosis methods, waiting times, and follow-up services.

In summary, affordability and accuracy are the two principal factors that should be considered when promoting AI-based health care. The combination of AI-based and professional health care will be more easily accepted by the general public as AI technology develops.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey introduction.
[DOCX File , 15 KB - jmir_v23i3e26997_app1.docx]

Multimedia Appendix 2
Supplementary questionnaire.
[DOCX File , 75 KB - jmir_v23i3e26997_app2.docx]

Multimedia Appendix 3
Propensity score matching method.
[DOCX File , 18 KB - jmir_v23i3e26997_app3.docx]

Multimedia Appendix 4
Random utility model.
[DOCX File , 18 KB - jmir_v23i3e26997_app4.docx]

Multimedia Appendix 5
Supplementary tables.
[DOCX File , 35 KB - jmir_v23i3e26997_app5.docx]

References

1. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial intelligence in health care: Bibliometric analysis. J Med Internet Res 2020 Jul 29;22(7):e18228 [FREE Full text] [doi: 10.2196/18228] [Medline: 32723713]
2. Vaira L, Bochicchio MA, Conte M, Casaluci FM, Melpignano A. MamaBot: a system based on ML and NLP for supporting women and families during pregnancy. In: IDEAS 2018: Proceedings of the 22nd International Database Engineering &

Applications Symposium. 2018 Jun Presented at: IDEAS 2018: 22nd International Database Engineering & Applications Symposium; June 2018; Villa San Giovanni, Italy p. 273-277. [doi: 10.1145/3216122.3216173]
3. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Jun 21;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
4. WHO coronavirus disease (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/table [accessed 2021-02-09]
5. AI-based COVID-19 CT report becomes a national collection. Alibaba DAMO Academy. 2020. URL: https://damo.alibaba.com/events/67 [accessed 2021-02-09]
6. Yassine HM, Shah Z. How could artificial intelligence aid in the fight against coronavirus? Expert Rev Anti Infect Ther 2020 Jun;18(6):493-497. [doi: 10.1080/14787210.2020.1744275] [Medline: 32223349]
7. Zandieh SO, Yoon-Flannery K, Kuperman GJ, Langsam DJ, Hyman D, Kaushal R. Challenges to EHR implementation in electronic- versus paper-based office practices. J Gen Intern Med 2008 Jun;23(6):755-761 [FREE Full text] [doi: 10.1007/s11606-008-0573-5] [Medline: 18369679]
8. From $600 M to $6 billion, artificial intelligence systems poised for dramatic market expansion in healthcare. Frost & Sullivan. 2016 Jan 05. URL: https://ww2.frost.com/news/press-releases/600-m-6-billion-artificial- [accessed 2021-02-09]
9. McFadden D. Regression-based specification tests for the multinomial logit model. J Econom 1987;34(1-2):63-82. [doi: 10.1016/0304-4076(87)90067-4]
10. Anas A. Discrete choice theory, information theory and the multinomial logit and gravity models. Transportation Research Part B: Methodological 1983 Feb;17(1):13-23. [doi: 10.1016/0191-2615(83)90023-1]
11. Greene WH, Hensher DA. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B: Methodological 2003 Sep;37(8):681-698. [doi: 10.1016/s0191-2615(02)00046-2]
12. Berhane A, Enquselassie F. Patients' preferences for attributes related to health care services at hospitals in Amhara Region, northern Ethiopia: a discrete choice experiment. Patient Prefer Adherence 2015 Sep 10;9:1293-1301. [doi: 10.2147/PPA.S87928] [Medline: 26396501]
13. Alizadeh AA, Eftekhaari T, Mousavi SH, Orouji GA, Mohammadi F, Javidan GA, et al. Patient preferences for hospital quality in Bandar Abbas using a discrete choice experiment: 2010-2011. Life Sci J 2012 Oct 15;9:1882-1886.
14. Ryan M. Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. Soc Sci Med 1999 Feb;48(4):535-546. [doi: 10.1016/s0277-9536(98)00374-8] [Medline: 10075178]
15. Gunst RF, Mason RL. Fractional factorial design. Wiley Interdiscip Rev Comput Stat 2009 Aug 19;1(2):234-244. [doi: 10.1002/wics.27]
16. A user guide with case studies: How to conduct a discrete choice experiment for health workforce recruitment and retention in remote and rural areas. World Health Organization. URL: https://www.who.int/hrh/resources/DCE_UserGuide_WEB.pdf?ua=1 [accessed 2021-02-09]
17. Bailey RA. Balance, orthogonality and efficiency factors in factorial design. J R Stat Soc Series B Stat Methodol 1985 Jul;47(3):453-458. [doi: 10.1111/j.2517-6161.1985.tb01374.x]
18. Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. J Econ Surv 2008 Jan 31;22(1):31-72. [doi: 10.1111/j.1467-6419.2007.00527.x]
19. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol 1974;66(5):688-701. [doi: 10.1037/h0037350]
20. Roy AD. Some thoughts on the distribution of earnings. Oxf Econ Pap 1951 Jun;3(2):135-146. [doi: 10.1093/oxfordjournals.oep.a041827]
21. Castaño-Muñoz J, Duart JM, Sancho-Vinuesa T. The internet in face-to-face higher education: Can interactive learning improve academic achievement? Br J Educ Technol 2013 Jan 14;45(1):149-159. [doi: 10.1111/bjet.12007]
22. Cost S, Salzberg S. A weighted nearest neighbor algorithm for learning with symbolic features. Mach Learn 1993 Jan;10:57-78. [doi: 10.1007/bf00993481]
23. Austin PC. Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. Stat Med 2016 Dec 30;35(30):5642-5655 [FREE Full text] [doi: 10.1002/sim.7084] [Medline: 27549016]
24. Hensher D, Greene W. The mixed logit model: The state of practice. Transportation (Amst) 2003;30:133-176. [doi: 10.1023/A:1022558715350]
25. Sakamoto Y, Ishiguro M, Kitagawa G, Reidel D. Akaike information criterion statistics. J Am Stat Assoc 1988 Sep;83(403):907. [doi: 10.2307/2289329]
26. Weakliem DL. A critique of the bayesian information criterion for model selection. Sociological Methods & Research 1999 Feb 01;27(3):359-397. [doi: 10.1177/0049124199027003002]
27. Hu R, Dong S, Zhao Y, Hu H, Li Z. Assessing potential spatial accessibility of health services in rural China: a case study of Donghai County. Int J Equity Health 2013 May 20;12:35 [FREE Full text] [doi: 10.1186/1475-9276-12-35] [Medline: 23688278]
28. Wang X, Yang H, Duan Z, Pan J. Spatial accessibility of primary health care in China: A case study in Sichuan Province. Soc Sci Med 2018 Jul;209:14-24. [doi: 10.1016/j.socscimed.2018.05.023] [Medline: 29778934]

29. Guo J, Li B. The application of medical artificial intelligence technology in rural areas of developing countries. Health Equity 2018;2(1):174-181 [FREE Full text] [doi: 10.1089/heq.2018.0037] [Medline: 30283865]

Abbreviations

AI: artificial intelligence
DCE: discrete choice experiment
LCM: latent class model
MNL: multinomial logit
OR: odds ratio
PSM: propensity score matching
WTP: willingness to pay

Edited by C Basch; submitted 07.01.21; peer-reviewed by H Imeri, B Sapkota; comments to author 28.01.21; revised version received 28.01.21; accepted 01.02.21; published 02.03.21.

Please cite as:
Liu T, Tsang W, Xie Y, Tian K, Huang F, Chen Y, Lau O, Feng G, Du J, Chu B, Shi T, Zhao J, Cai Y, Hu X, Akinwunmi B, Huang J, Zhang CJP, Ming WK
Preferences for Artificial Intelligence Clinicians Before and During the COVID-19 Pandemic: Discrete Choice Experiment and Propensity Score Matching Study
J Med Internet Res 2021;23(3):e26997
URL: https://www.jmir.org/2021/3/e26997
doi:10.2196/26997
PMID:33556034

Original Paper

Artificial Intelligence Techniques That May Be Applied to Primary Care Data to Facilitate Earlier Diagnosis of Cancer: Systematic Review

Owain T Jones[1], MPhil; Natalia Calanzani[1], PhD; Smiji Saji[1], MBBCHIR; Stephen W Duffy[2], MSc; Jon Emery[3], DPhil; Willie Hamilton[4], MD; Hardeep Singh[5], MD, MPH; Niek J de Wit[6], MD; Fiona M Walter[1], MD

[1]Primary Care Unit, Department of Public Health & Primary Care, University of Cambridge, Cambridge, United Kingdom

[2]Wolfson Institute for Preventive Medicine, Queen Mary University of London, London, United Kingdom

[3]Centre for Cancer Research and Department of General Practice, University of Melbourne, Victoria, Australia

[4]College of Medicine and Health, University of Exeter, Exeter, United Kingdom

[5]Center for Innovations in Quality, Effectiveness and Safety, Michael E DeBakey Veterans Affairs Medical Center and Baylor College of Medicine, Houston, TX, United States

[6]Julius Center for Health Sciences and Primary Care, UMC Utrecht, Utrecht, Netherlands

Corresponding Author:
Owain T Jones, MPhil
Primary Care Unit
Department of Public Health & Primary Care
University of Cambridge
2 Wort's Causeway
Cambridge, CB1 8RN
United Kingdom
Phone: 44 1223762554
Email: otj24@medschl.cam.ac.uk

Abstract

Background: More than 17 million people worldwide, including 360,000 people in the United Kingdom, were diagnosed with cancer in 2018. Cancer prognosis and disease burden are highly dependent on the disease stage at diagnosis. Most people diagnosed with cancer first present in primary care settings, where improved assessment of the (often vague) presenting symptoms of cancer could lead to earlier detection and improved outcomes for patients. There is accumulating evidence that artificial intelligence (AI) can assist clinicians in making better clinical decisions in some areas of health care.

Objective: This study aimed to systematically review AI techniques that may facilitate earlier diagnosis of cancer and could be applied to primary care electronic health record (EHR) data. The quality of the evidence, the phase of development the AI techniques have reached, the gaps that exist in the evidence, and the potential for use in primary care were evaluated.

Methods: We searched MEDLINE, Embase, SCOPUS, and Web of Science databases from January 01, 2000, to June 11, 2019, and included all studies providing evidence for the accuracy or effectiveness of applying AI techniques for the early detection of cancer, which may be applicable to primary care EHRs. We included all study designs in all settings and languages. These searches were extended through a scoping review of AI-based commercial technologies. The main outcomes assessed were measures of diagnostic accuracy for cancer.

Results: We identified 10,456 studies; 16 studies met the inclusion criteria, representing the data of 3,862,910 patients. A total of 13 studies described the initial development and testing of AI algorithms, and 3 studies described the validation of an AI algorithm in independent data sets. One study was based on prospectively collected data; only 3 studies were based on primary care data. We found no data on implementation barriers or cost-effectiveness. Risk of bias assessment highlighted a wide range of study quality. The additional scoping review of commercial AI technologies identified 21 technologies, only 1 meeting our inclusion criteria. Meta-analysis was not undertaken because of the heterogeneity of AI modalities, data set characteristics, and outcome measures.

Conclusions: AI techniques have been applied to EHR-type data to facilitate early diagnosis of cancer, but their use in primary care settings is still at an early stage of maturity. Further evidence is needed on their performance using primary care data,

implementation barriers, and cost-effectiveness before widespread adoption into routine primary care clinical practice can be recommended.

(J Med Internet Res 2021;23(3):e23483) doi:10.2196/23483

KEYWORDS

artificial intelligence; machine learning; electronic health records; primary health care; early detection of cancer

Introduction

Background

Cancer control is a global health priority, with 17 million new cases diagnosed worldwide in 2018. In high-income countries such as the United Kingdom, approximately half the population over the age of 50 years will be diagnosed with cancer in their lifetime [1]. Although the National Health Service (NHS) currently spends approximately £1 billion (US $1.37 billion) on cancer diagnostics per year [2], the United Kingdom lags behind comparable European nations with their cancer survival rates [3].

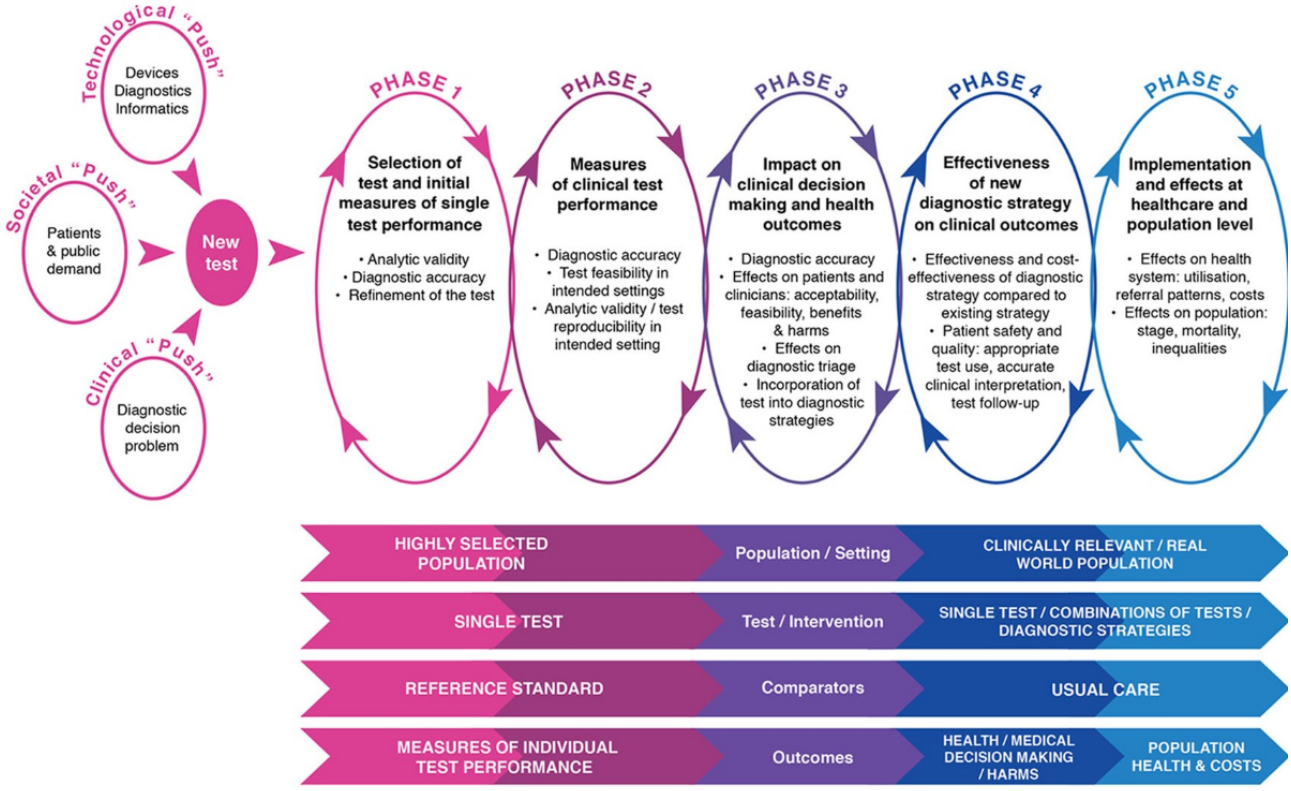
In gatekeeper health care systems such as the United Kingdom, most people diagnosed with cancer first present in primary care [4], where general practitioners evaluate (often vague) presenting symptoms and decide on an appropriate management strategy, including investigations, specialist referral, or reassurance. More accurate assessment of these symptoms, especially for patients with multiple consultations, could lead to earlier diagnosis of cancer and improved outcomes for patients, including improved survival rates [5,6].

There is accumulating evidence that artificial intelligence (AI) can assist clinicians in making better clinical decisions or even replace human judgment, in certain areas of health care. This is due to the increasing availability of health care data and the rapid development of big data analytic methods. There has been increasing interest in the application of AI in medical diagnosis, including machine learning and automated analysis approaches. Recent studies have applied AI to patient symptoms to improve diagnosis [7,8], to retinal images for the diagnosis of diabetic retinopathy [9], to mammography images for breast cancer diagnosis [10,11], to computed tomography (CT) scans for the diagnosis of intracranial hemorrhages [12], and to images of blood films for the diagnosis of acute lymphoblastic leukemia [13].

Few AI techniques are currently implemented in routine clinical care. This may be due to uncertainty over the suitability of current regulations to assess the safety and efficacy of AI systems [14-16], a lack of evidence about the cost-effectiveness and acceptability of AI systems [14], challenges to implementation into existing electronic health records (EHRs) and routine clinical care, and uncertainty over the ethics of using AI systems. A recent review of AI and primary care reported that research on AI for primary care is at an early stage of maturity [17], although research on AI-driven tools such as symptom checkers for patient and clinical users are more mature [18-21].

The CanTest framework [22] (Figure 1) establishes the developmental phases required to ensure that new diagnostic tests or technologies are fit for purpose when introduced into clinical practice. It provides a roadmap for developers and policy makers to bridge the gap from the development of a diagnostic test or technology to its successful implementation. We used this framework to guide the assessment of the studies identified in this review.

Figure 1. The CanTest Framework [22].



Objectives

Few studies of AI-based techniques for the early detection of cancer have been undertaken in primary care settings [17]. Therefore, the aim of this systematic review is to identify AI techniques that facilitate the early detection of cancer and could be applied to primary care EHR data. We also aim to summarize the diagnostic accuracy measures used to evaluate existing studies and evaluate the quality of the evidence, the phase of development the AI technologies have reached, the gaps that exist in the evidence, and the potential for use in primary care. As many commercial technological developments are not documented in academic publications, we also performed a parallel scoping review of commercially available AI-based technologies for the early detection of cancer that may be suitable for implementation in primary care settings.

Methods

Search Strategy and Selection Criteria

This study was conducted in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis) guidelines [23], and the protocol was registered with PROSPERO (an international prospective register of systematic reviews) before conducting the review (CRD42020176674) [24]. All aspects of the protocol were reviewed by the senior research team.

We included all primary research articles published in peer-reviewed journals, without language restrictions, from January 01, 2000, to June 11, 2019. Studies were included if they provided evidence around the accuracy, utility, acceptability, or cost-effectiveness of applying AI techniques to facilitate the early detection of cancer and could be applied to primary care EHRs (ie, to the types of data found in primary care EHRs) [22]. We included AI techniques based on any type of data that were relevant to primary care settings, including coded data and free text. We included all types of study design, as we anticipated that there would be few relevant randomized controlled trials. We kept our search terms broad to not miss relevant studies and carefully considered evidence from any health care system to assess whether the evidence could be applied to primary care settings.

As our aim is to identify AI techniques that would be applicable in primary care clinical settings, we excluded studies that incorporated data not typically available in primary care EHRs in the early diagnostic stages (eg, histopathology images, magnetic resonance imaging, or CT scan images). We also excluded studies that only described the development of an AI technique without any testing or evaluation data, studies that did not incorporate an element of machine learning (ie, with training and testing or validation steps), studies that used AI techniques for biomarker discovery alone, and studies that were based on sample sizes of less than 50 cases or controls. Machine learning techniques and neural networks have been described since the 1960s [25,26]; however, they were initially limited by computing power and data availability. We chose to start our search in 2000, as this was when the earliest research describing the new wave of machine learning techniques emerged [27].

We searched MEDLINE, Embase, SCOPUS, and Web of Science bibliographic databases, using keywords related to AI, cancer, and early detection. We extended these systematic

searches through manual searching of the reference lists of the included studies. We contacted study authors, where required. Where studies were not published in English, we identified suitably qualified native speakers to help assess these studies. We performed a parallel scoping review to look for commercially developed AI technologies that were not identified through systematic searches, thus unpublished and not scientifically evaluated. This included manually searching commercial research archives and networks (eg, arXiv [28], Google [29], Microsoft [30], and IBM [31]), reviewing the computer-based technologies identified in 3 recent reviews [19-21], and manually searching for further technologies mentioned in the text or references of the studies and websites included in these reviews.

Following duplicate removal, 1 author (OJ) screened titles and abstracts to identify studies that fit the inclusion criteria. Of the titles and abstracts, 17.42% (1838/10,456) were checked by 2 other authors (SS and NC); interrater reliability was excellent at 96.24% (1769/1838). Any disagreements were discussed by the core research team (OJ, SS, NC, and FW), and a consensus was reached. Three reviewers (OJ, SS, and NC) independently assessed the full-text articles for inclusion in the review. Any disagreements were resolved by a consensus-based decision.

Data Analysis

Data extraction was undertaken independently by at least two reviewers (OJ, SS, and NC) into a predesigned data extraction spreadsheet. The research team met regularly to reach consensus by discussing and resolving any differences in data extraction. One author (OJ) amalgamated the data extraction spreadsheets, summarizing the data where possible.

The main summary measures collected included sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), area under the receiver operating characteristic (AUROC) curve, and any other diagnostic accuracy measures of the AI techniques. Secondary outcomes include the types of AI used, the type of data used to train and test the algorithms, and how these algorithms were evaluated. We also collected data, where identified, on cost-effectiveness and patient or clinician acceptability.

Risk of bias assessment was undertaken for all full-text papers by 2 independent researchers (OJ and NC) using the quality assessment of diagnostic accuracy studies-2 (QUADAS-2) critical appraisal tool [32]. OJ assessed all studies, and 50% (40/79) of them were cross-checked by NC. Any disagreements in the assessment were resolved by consensus discussion.

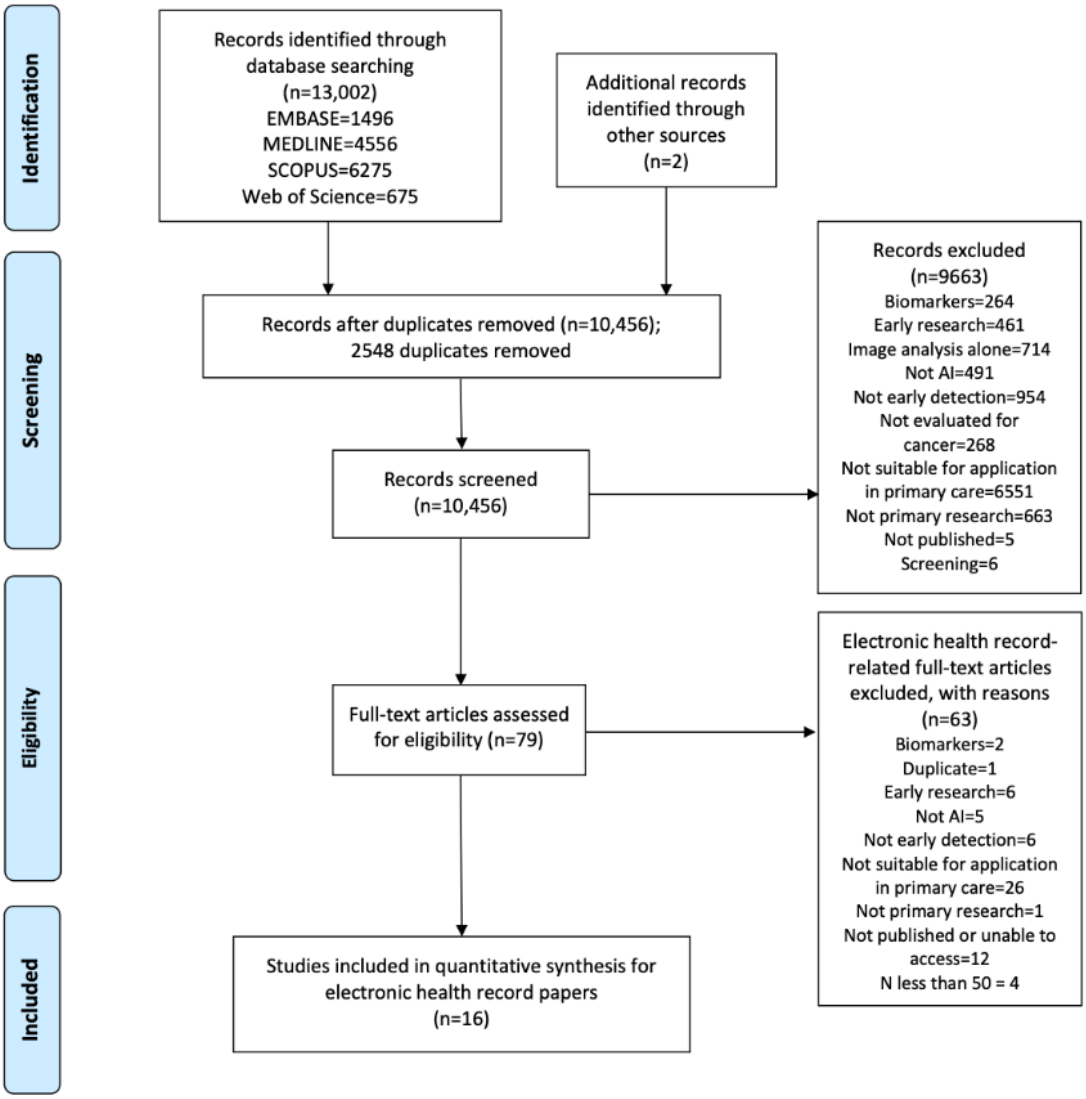
The studies identified were heterogeneous, employing various AI techniques and using different outcome measures for evaluation. Hence, a meta-analysis of the data was not possible, and we chose to use a narrative synthesis approach, following established guidance on its methodology [33]. We aimed to summarize the findings of the identified studies using primarily a textual approach, while also providing an overview of the quantitative outcome measures used in the studies. Once data extraction was completed, we explored the relationships that emerged within the data.

Full details of our review question, search strategy, inclusion or exclusion criteria, and data extraction methodology are described in Multimedia Appendices 1 [1-5,7-9,11-13,34-38] and 2, and the full list of excluded studies is provided in Multimedia Appendix 3 [34,39-114].

Results

A total of 13,004 articles were identified in database searches (including 2548 duplicates), and 793 articles underwent full-text review. Of the 79 articles that were related to EHRs, 16 met the inclusion criteria and were included in this analysis (Figure 2), representing the data of 3,862,910 patients. No articles identified through other sources or reference lists met the inclusion criteria.

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis) flow diagram for studies included in the review. AI: artificial intelligence.



Tables 1 and 2 show the main study characteristics for the 16 included studies, including the modality of AI used. Supplementary information on the variables included in the AI techniques is available in Multimedia Appendix 4 [34,39-53]. We categorized the variables included into the following categories: demographics, symptoms, comorbidities, lifestyle history, examination findings, blood results, and other. Most studies (n=13) described the initial development and testing of an AI technique [39-51]. Three studies validated the AI technique developed by Kinar et al [48] in independent data sets from 3 different countries (Israel, United States, and United Kingdom) [34,52,53].

Table 1. Study details including modality of artificial intelligence and adopted comparison or control.

Study	Authors' origin	Cancer	Modality of artificial intelligence	Comparison or control			
				Histopathology	Specialist	Not stated	Other
Development studies							
Alzubi et al, 2019 [39]	Jordan and India	Lung cancer	WONN-MLB[a]	X[b]	—[c]	—	1[d]
Chang et al, 2009 [40]	Taiwan	Pancreatic Cancer	BPNN[e]; LR[f]	—	—	X	2[g], 3[h]
Cooper et al, 2018 [41]	United Kingdom	Colorectal Cancer	ANN[i]; CVT[j]; LR	X	X	—	4[k]
Cowley et al, 2013 [42]	United Kingdom	Colorectal Cancer	BPANN[l]	—	X	—	2; 5[m]
Daqqa et al, 2017 [43]	Gaza, Palestine	Leukemia	SVM[n]; DT[o]; K-NN[p]	X	—	—	2
Goryński et al, 2014 [44]	Poland	Lung cancer	MLP-ANN[q]	X	X	—	—
Hart et al, 2018 [45]	United States	Lung cancer	BPANN	—	—	X	2; 6[r]
Kalra et al, 2003 [46]	United States	Prostate cancer	BPNN	X	—	—	2; 3
Kang et al, 2017 [47]	China	Any cancer	BPNN; CVT; SVM; DT	X	X	—	2
Kinar et al, 2016 [48]	Israel and United States	Colorectal Cancer	DT/RF[s]; GBM[t]; CVT	X	X	—	3; 6
Kop et al, 2016 [49]	The Netherlands	Colorectal Cancer	CART[u]; RF; LR; CVT	X	X	—	—
Miotto et al, 2016 [50]	United States	Multiple diseases and cancers	DNN[v]; RF	—	X	—	2; 3
Payandeh et al, 2009 [51]	Iran	CML[w] and lymphoproliferative disorders	MLP-ANN	X	X	—	3
Validation studies							
Birks et al, 2017 [52]	United Kingdom	Colorectal Cancer	DT/RF; GBM; CVT	X	X	—	—
Hornbrook et al, 2017 [34]	United States	Colorectal Cancer	DT/RF; GBM; CVT	X	X	—	—
Kinar et al, 2017 [53]	Israel	Colorectal Cancer	DT/RF; GBM; CVT	X	X	—	—

[a]WONN-MLB: weight optimized neural network with maximum likelihood boosting.

[b]X: corresponding control used in this study.

[c]Not used in this study.

[d]1: previously developed artificial intelligence methods.

[e]BPNN: back propagation neural network.

[f]LR: logistic regression.

[g]2: other artificial intelligence methods developed by this author.

[h]3: other statistical (ie, non-artificial intelligence) techniques.

[i]ANN: artificial neural network.

[j]CVT: cross-validation techniques.

[k]4: colonoscopy.

[l]BPANN: back propagation artificial neural network.

[m]5: primary care clinicians.

[n]SVM: support vector machine.

[o]DT: decision tree.

[p]K-NN: K-nearest neighbor.

[q]MLP-ANN: multilayer perceptron artificial neural network.

[r]6: screening tests (eg, low-dose computed tomography scan and fecal occult blood test).

[s]RF: random forest.

[t]GBM: gradient boosting model.

[u]CART: classification and regression trees.

[v]DNN: deep neural network.

[w]CML: chronic myeloid leukemia.

The study authors originated from a variety of countries, including the United States (n=5), countries in the Middle East (n=5), Europe (n=5), and Asia (n=3), with some studies involving multiple countries. The AI techniques were most commonly developed to identify colorectal cancer (n=7) [34,41,42,48,49,52,53], although they also addressed lung cancer (n=3) [39,44,45], hematological cancers (n=2) [43,51], pancreatic cancer (n=1) [40], prostate cancer (n=1) [46], and multiple cancers (n=2) [47,50].

Neural networks were the dominant technique employed (n=10) [39-42,44-47,50,51], with many neural network subtypes mentioned. The study by Miotto et al [50] was the only study to include a processed form of the free text notes in the data used by the AI technique, although the work described by Kop et al [49] was developed in a subsequent study to include clinical free text data [115].

The majority of studies (n=9) used a combination of histopathological diagnoses and expert opinion as the control for their study [34,41,44,47-49,51-53]. The clinical control group was unclear in 2 studies [40,45]. Many studies used multiple AI techniques and then compared them with each other (n=8) [40,42,43,45-47,49,50]. Some studies used non-AI techniques, such as logistic regression and screening tests, as comparators for the performance of the AI technique that was being developed [40,41,45,46,48-51].

Table 2. Study details: patient variables.

Study	Patient variables									
	Age	Sex	Demographics	Symptoms	Comorbidities	Lifestyle	Examination	FBC[a]	Other blood tests	Other[b]
Development studies										
Alzubi et al, 2019 [39]	X[c]	—[d]	—	X	X	X	—	—	—	X
Chang et al, 2009 [40]	X	X	—	X	X	X	—	X	X	—
Cooper et al, 2018 [41]	X	X	X	—	—	—	—	—	—	X
Cowley et al, 2013 [42]	—	—	—	X	X	X	—	—	—	X
Daqqa et al, 2017 [43]	—	—	—	—	—	—	—	X	—	—
Goryński et al, 2014 [44]	X	X	X	X	X	X	X	X	X	X
Hart et al, 2018 [45]	X	X	X	—	X	X	X	—	—	—
Kalra et al, 2003 [46]	X	—	X	X	X	—	X	—	X	—
Kang et al, 2017 [47]	X	X	—	—	—	—	X	X	X	X
Kinar et al, 2016 [48]	X	X	—	—	—	—	—	X	—	—
Kop et al, 2016 [49]	X	X	—	X	X	X	X	X	X	X
Miotto et al, 2016 [50]	—	—	X	X	X	X	X	—	X	X
Payandeh et al, 2009 [51]	—	—	—	—	—	—	—	X	—	—
Validation studies										
Birks et al, 2017 [52]	X	X	—	—	—	—	—	X	—	—
Hornbrook et al, 2017 [34]	X	X	—	—	—	—	—	X	—	—
Kinar et al, 2017 [53]	X	X	—	—	—	—	—	X	—	—

[a]FBC: full blood count.

[b]More detail on other variables included is available in Multimedia Appendix 4.

[c]X: corresponding variable used in this study.

[d]Not used in this study.

Most of the studies (n=12) included blood test results, all suitable for use in primary care settings. Age was also commonly included (n=12). Other variables used were sex (n=10), demographics (n=5), symptoms (n=7), comorbidities (n=8), lifestyle history (n=7), examination findings (n=6), medication or prescription history (n=3), spirometry results (n=2), urine dipstick results (n=1), fecal immunochemical test results (n=1), x-ray text reports (n=1), and referrals (n=1).

Table 3 shows the study designs and populations. Most studies used data sets originating from specialist care settings (n=7) [39,40,42-44,46,51], with only 3 studies using solely primary care patient data [41,49,52]. Kinar et al [48] included a follow-up validation study based on the health improvement network (THIN) database, also using primary care data. Several studies used a mixture of primary and secondary care patient data (n=5) [34,47,48,50,53].

Table 3. Study population and study design.

Study details	Population from health care setting	Database used	Disease positive population (patients)	Disease negative population (patients)	Training set (patients)	Testing set (patients)
Development studies						
Alzubi et al, 2019 [39]	Specialist care	Wroclaw Thoracic Surgery Centre	1200 in total; numbers of disease positive and negative unclear	1200 in total; numbers of disease positive and negative unclear	N/S[a]	1000
Chang et al, 2009 [40]	Specialist care (unclear)	"a certain medical center"	194	157[b]	234	117
Cooper et al, 2018 [41]	Primary care	NHS[c] Bowel Cancer Screening Programme comparative study [116]	549	1261	N/S	N/S
Cowley et al, 2013 [42]	Specialist care	2-week wait colorectal referrals to Castle Hill Hospital	74	703	777	100
Daqqa et al, 2017 [43]	Specialist care	Complete Blood Count test repository, European Gaza Hospital	2000	2000	N/S	N/S
Goryński et al, 2014 [44]	Specialist care	Patients treated at Kuyavia and Pomerania Centre of pulmonology	103	90	97	48
Hart et al, 2018 [45]	Other (survey)	National Health Interview Survey	649	488,418	342,347	146,719
Kalra et al, 2003 [46]	Specialist care	Men whose samples were tested at 6 sites in the United States[d]	348	N/S	218	144
Kang et al, 2017 [47]	Mixed	Database of Ci Ming Health Checkup Center	650	1650	N/S	N/S
Kinar et al, 2016 [48][e]	Mixed	Maccabi Health Services EMRs[f] linked to the Israel Cancer Registry	2437	463,670	466,107	139,205
Kop et al, 2016 [49]	Primary care	6 anonymized data sets from 3 urban regions, each covering a GP[g] recording system	1292	263,879	N/S	N/S
Miotto et al, 2016 [50]	Mixed	Mount Sinai Data Warehouse	276,214 patients with 78 diseases	276,214 patients with 78 diseases	200,000	76,214
Payandeh et al, 2009 [51]	Specialist care	Blood test results from patients at the Taleghani Hospital	450	N/S	360	132
Validation studies						
Birks J et al, 2017 [52]	Primary care	Clinical Practice Research Datalink	5141	2,220,108	N/A[h]	N/A
Hornbrook et al, 2017 [34]	Mixed	Kaiser Permanente North West EHR[i] system, Kaiser Permanente Tumor Registry	900	16,195	N/A	N/A
Kinar et al, 2017 [53]	Mixed	Maccabi Health Services EMRs, linked to the Israel Cancer Registry	133	112,451	N/A	N/A

[a]N/S: not stated.

[b]Cases of acute pancreatitis.

[c]NHS: National Health Service.

[d]Hospitals included: Northwest Prostate Institute Seattle, the University of Washington Seattle, the Johns Hopkins Hospital Baltimore, Memorial Sloan-Kettering Cancer Institute New York, Brigham and Women's Hospital Boston, and The University of Texas MD Anderson Cancer Center

[e]NB: this study also included a small validation study in the Health Improvement Network database in the United Kingdom (n=25,613)

[f]EMR: electronic medical record.

[g]GP: general practitioner.

[h]N/A: not applicable

[i]EHR: electronic health record.

Almost all the studies used different data sets, with the exception of the Maccabi Health Services EHR, which was used in 2 studies [48,53]. The data set sizes ranged from 193 to 2,225,249 patients, with a mean of 241,585 (SD 555,953), median of 3,150, and IQR of 267,237 patients. The wide range is primarily due to the large data set used by Birks et al [52]. Of the 13 development studies, 3 provided no information on the control population used [39,46,51]. Five of the development studies did not provide full information on how they partitioned their data set for the training and testing of the algorithm [39,41,43,47,49]. Five studies appeared to have independent training and testing data sets, with most split in ratios ranging from 60:40 to 70:30 [40,44-46,50].

Three studies [34,52,53] validated a previously developed AI technique [48] in independent data sets. Kinar et al [48] reported both the initial development of an AI technique and a subsequent validation study in an independent data set. The study by Cooper et al [41] was the only study that developed an AI technique based on prospectively collected clinical data, with the data originating from a pilot study of fecal immunochemical testing by the NHS Bowel Cancer Screening Programme [116].

Table 4 summarizes the main reported outcome measures. Specificity (n=11), AUROC (n=11), and sensitivity (n=10) were the most frequently reported; others included PPV (n=6), NPV (n=5), diagnostic accuracy (n=4), and odds ratios (n=3). Specificity results range from 80.6% [45] to 100% [51], sensitivity results from 0% [51] to 96.7% [40], and AUROC results from 0.55 [45] to 0.9896 [44].

Table 4. Outcome measures.

Study	Cancer type	Outcome measures for each modality of AI[a]
Development studies		
Alzubi et al, 2019 [39]	Lung cancer	• Specificity: 92%, Accuracy: 93% • False positive rate: 9%, F-1 score: 92%
Chang et al, 2009 [40]	Pancreatic cancer	• Sensitivity: BPNN[b] 88.3%, genetic algorithm LR[c] 96.7%, stepwise LR 96.7% • Specificity: BPNN 84.2%, genetic algorithm LR 82.5%, stepwise LR 73.7% • AUROC[d]: BPNN 0.895, genetic algorithm LR 0.921, stepwise LR 0.882
Cooper et al, 2018 [41]	Colorectal cancer	• Sensitivity: 35.15% (at FIT[e] threshold 160 μg g^{-1}) • Specificity: 85.57% • PPV[f]: 51.47%, NPV[g]: 75.19%, AUROC: 0.69, cancer detection rate: 10.66%
Cowley et al, 2013 [42]	Colorectal cancer	• Sensitivity: 90% • Specificity: 96% • PPV: 62%, NPV: 99%
Daqqa et al, 2017 [43]	Leukemia	• Sensitivity: SVM[h] 69.7%, K-NN[i] 60.0%, decision tree 62.4% • Specificity: SVM 81.5%, K-NN 82.8%, decision tree 87.1% • PPV: SVM 71.3%, K-NN 68.1%, decision tree 76.1% • NPV: SVM 80.4%, K-NN 74.1%, decision tree 87.1% • Accuracy: SVM 76.82%, K-NN 72.15%, decision tree 77.3% • F-measure: SVM 70%, K-NN 60%, decision tree 67%
Goryński et al, 2014 [44]	Lung cancer	• AUROC: 0.9896
Hart et al, 2018 [45]	Lung cancer	• Sensitivity: ANN[j] 75.30% • Specificity: ANN 80.60% • AUROC: ANN 0.86, RF[k] 0.81, SVM 0.55
Kalra et al, 2003 [46]	Prostate cancer	• Specificity: 92% • AUROC: 0.825
Kang et al, 2017 [47]	Any cancer	• Sensitivity: DNN[l] 64.07%, SVM 54.46%, decision tree 60.00% • Specificity: DNN 94.77%, SVM 95.27%, decision tree 91.50% • AUROC: DNN 0.882, SVM 0.928, decision tree 0.824 • Accuracy: DNN 86.00%, SVM 83.83%, decision tree 83.60% • Using fuzzy interval of threshold with DNN achieves sensitivity 90.20%, specificity 94.22%, accuracy 93.22%
Kinar et al, 2016 [48]	Colorectal cancer	• Specificity: Testing set 88% overall (at a sensitivity of 50%). Higher for proximal colon tumors. Validation set 94% (at a sensitivity of 50%) • AUROC: Testing set 0.82, validation set 0.81 • OR[m] 26 at false +ve rate of 0.5% (testing set), OR 40 at false +ve rate of 0.5% (validation set). Algorithm identified 48% more CRC[n] cases than gFOBT[o]
Kop et al, 2016 [49]	Colorectal cancer	• Sensitivity: CART[p] 53.9%, RF 63.7%, LR 64.2% • PPV: CART 2.6%, RF 3%, LR 3% • AUROC: CART 0.885, RF 0.889, LR 0.891 • F1-score: CART 0.049, RF 0.057, LR 0.058. • Drugs for constipation most important predictor of CRC, followed by iron deficiency anemia

Study	Cancer type	Outcome measures for each modality of AI[a]
Miotto et al, 2016 [50]	Multiple diseases and cancers	• Specificity: 92% • AUROC: 0.773 for classification of all diseases (cancer and other diagnoses). Rectal or anal cancer 0.887, liver or intrahepatic bile duct cancer 0.886, prostate cancer 0.859, multiple myeloma 0.849, ovarian cancer 0.824, bladder cancer 0.818, testicular cancer 0.811, pancreatic cancer 0.795, leukemia 0.774, uterine cancer 0.771, non-Hodgkin lymphoma 0.771, bronchial or lung cancer 0.770, colon cancer 0.767, breast cancer 0.762, kidney or renal pelvis cancer 0.753, brain or nervous system cancer 0.742, Hodgkin disease 0.731, cervical cancer 0.675 • Accuracy index: 0.929 overall for classification of all diseases • F-score: 0.181 for classification of all diseases • Deep patient obtained approximately 55% correct predictions when suggesting 3 or more diseases per patient, regardless of time interval
Payandeh et al, 2009 [51]	CML[q] and lymphopro-liferative disorders	• Sensitivity: CML 0%, lymphoproliferative disorder 0% • Specificity: CML 100%, lymphoproliferative disorder 99.2% • PPV: CML 0%, lymphoproliferative disorder 0% • NPV: CML 99.2%, lymphoproliferative disorder 100% • Error % for convoluted neural network 0.33, error % for LR 0.78
Validation studies		
Birks et al, 2017 [52]	Colorectal cancer	• AUROC: analyzed at various time intervals before diagnosis, 3-6 months 0.844, 18-24 months 0.776
Hornbrook et al, 2017 [34]	Colorectal cancer	• Sensitivity: 0-180 days (test to diagnosis): 50-75 years: 34.5%, 40-89 years: 39.9%; 181-360 days: 50-75 years: 18.8%, 40-89 years: 27.4% • AUROC: 0.80, OR: 34.7 at 99% specificity, 19.7 at 97%, 14.6 at 95%, 10.0 at 90%
Kinar et al, 2017 [53]	Colorectal cancer	• Sensitivity: 17.0% at 1% +ve rate, 24.4% at 3% +ve rate • PPV: 2.1% at 1% +ve rate, 1.0% at 3% +ve rate • NPV: 99.9% at 1% +ve rate, 99.9% at 3% +ve rate • OR: 21.8% at 1% +ve rate, 10.9% at 3% +ve rate

[a]AI: artificial intelligence.
[b]BPNN: back propagation neural network.
[c]LR: logistic regression.
[d]AUROC: area under the receiver operating characteristic.
[e]FIT: fecal immunochemical test.
[f]PPV: positive predictive value.
[g]NPV: negative predictive value.
[h]SVM: support vector machine.
[i]K-NN: K-nearest neighbor.
[j]ANN: artificial neural network.
[k]RF: random forest.
[l]DNN: deep neural network.
[m]OR: odds ratio.
[n]CRC: colorectal cancer.
[o]gFOBT: guaiac fecal occult blood test.
[p]CART: classification and regression trees.
[q]CML: chronic myeloid leukemia.

We looked for other secondary outcomes, including implementation barriers to AI techniques in primary care settings, but did not find any evidence related to patient or clinician acceptability or cost-effectiveness.

Table 5 shows the outcomes of the risk of bias assessment using the QUADAS-2 tool. The studies demonstrated a wide range in quality; however, no studies were excluded based on their risk of bias assessment. The identified limitations were acknowledged in the relative contribution of the studies to the conclusions of the review.

Table 5. Critical appraisal results using the Quality Assessment of Diagnostic Accuracy Studies-2 tool.

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Alzubi et al, 2019 [39]	✗ [a]	✗ [b]	✗	✗ [c]	✗	✗	✗
Birks et al, 2017 [52]	✗	✗	✗	✗	✗	✗	✗
Chang et al, 2009 [40]	✗	✗	✗	✗	✗	✗	✗
Cooper et al, 2018 [41]	✗	✗	✗	✗	✗	✗	✗
Cowley et al, 2013 [42]	✗	✗	✗	✗	✗	✗	✗
Daqqa et al, 2017 [43]	✗	✗	✗	✗	✗	✗	✗
Goryński et al, 2014 [44]	✗	✗	✗	✗	✗	✗	✗
Hart et al, 2018 [45]	✗	✗	✗	✗	✗	✗	✗
Hornbrook et al, 2017 [34]	✗	✗	✗	✗	✗	✗	✗
Kalra et al, 2003 [46]	✗	✗	✗	✗	✗	✗	✗
Kang et al, 2017 [47]	✗	✗	✗	✗	✗	✗	✗
Kinar et al, 2016 [48]	✗	✗	✗	✗	✗	✗	✗
Kinar et al, 2017 [53]	✗	✗	✗	✗	✗	✗	✗
Kop et al, 2016 [49]	✗	✗	✗	✗	✗	✗	✗
Miotto et al, 2016 [50]	✗	✗	✗	✗	✗	✗	✗
Payandeh et al, 2009 [51]	✗	✗	✗	✗	✗	✗	✗

[a]High risk.

[b]Low risk.

[c]Unclear risk.

Table 6 summarizes the computer-based technologies identified in our parallel scoping review of commercial AI technologies. We identified 21 commercial computer-based technologies. Of these, 11 were clinician-facing differential diagnosis technologies that did not appear to be integrated into the EHR [117-127]. Ten of the technologies were linked to, or integrated into, the EHR in some way [8,128-136]. Nine of the technologies did not use AI algorithms incorporating an element of machine learning, as was required in our inclusion criteria [118,120-127]. It was also not clear from the websites and studies of 3 further technologies whether they met our AI inclusion criteria [117,130,134]. There were 8 technologies that met our inclusion criteria for AI (Abtrace [128], Babylon [8], Cthesigns [129], Isabel [131], Medial EarlySign [132], symcat [119], symptomate [135], and the unnamed technology evaluated by Liang et al [136]). Only the Medial EarlySign tool was evaluated for its performance in the diagnosis or triage of potential cancer [132]; 4 of the studies developing and validating this technology were included in this systematic review [34,48,52,53]. Cthesigns is specifically designed to aid the early diagnosis of cancer but has not been the subject of any studies we could identify [129].

Table 6. Summarizing scoping review of commercial artificial intelligence technologies.

Technology identified (origin) websites and associated academic studies	Not AI[a]	Not cancer	Not primary care based	Not early detection or diagnosis	Early research	Not published	Not primary research	<50 cases or controls
Abtrace (United Kingdom)								
Abtrace website [128]	—[b]	—	—	—	—	X[c]	—	—
Babylon (United Kingdom)								
Babylon health website [8]	—	—	—	—	—	—	—	—
Zhelezniak et al [137]	—	X	X	X	X	—	—	—
Douglas et al [138]	—	X	X	X	X	—	—	—
Smith et al [139]	—	X	X	X	X			
National Health Service 111 powered by Babylon - Outcomes Evaluation [140]	—	X	—	X	—	—	—	—
Middleton et al [141]	—	X	—	X	—	—	—	—
Cthesigns (United Kingdom)								
Cthesigns website [129]	—	—	—	—	—	X	—	—
Diagnosis Pro (United States)								
No website identified	—	—	—	—	—	—	—	—
Bond et al [117]	N/C[d]	X	—	—	—	—	—	—
DocResponse (United States)								
Docresponse website [130]	N/C	—	—	—	—	X	—	—
DxPlain (United States)								
Dxplain website [118]	N/C	—	—	—	—	—	—	—
Barnett et al [142]	X	—	—	—	X	—	X	—
Barnett et al [143]	X	X	—	—	—	—	—	—
Bauer et al [144]	X	X	—	—	—	—	—	—
Berner et al [145]	X	X	X	—	—	—	—	—
Bond et al [117]	X	X	—	—	—	—	—	X
Elhanan et al [146]	X	—	—	—	X	—	—	—
Elkin et al [147]	X	X	X	—	—	—	—	—
Feldman et al [148]	X	X	X	—	—	—	—	X
Hammersley et al [149]	X	X	X	—	—	—	—	—
Hoffer et al [150]	X	—	—	X	—	—	—	—
London et al [151]	X	—	—	—	X	—	—	—
Iliad (United States)								
No website identified	—	—	—	—	—	—	—	—
Berner et al [145]	X	X	X	—	—	—	—	—
Elstein et al [152]	X	X	X	—	—	—	—	X
Friedman et al [153]	X	—	X	—	—	—	—	X
Gozum et al [154]	X	—	X	—	—	—	—	X
Graber et al [155]	X	—	X	—	—	—	—	X
Heckerling et al [120]	X	—	X	—	—	—	—	X
Lange et al [156]	X	—	—	—	—	—	—	X
Lau et al [157]	—	—	—	—	—	—	X	—
Li et al [158]	X	X	X	—	—	—	—	X

Technology identified (origin) websites and associated academic studies	Not AI[a]	Not cancer	Not primary care based	Not early detection or diagnosis	Early research	Not published	Not primary research	<50 cases or controls
Lincoln et al [159]	X	X	X	—	—	—	—	X
Murphy et al [160]	X	X	X	—	—	—	—	X
Wolf et al [161]	X	X	X	—	—	—	—	X
Internist-1 (United States)								
No website identified	—	—	—	—	—	—	—	—
Miller et al [121]	X	X	X	—	—	—	—	X
Miller et al [122]	X	—	X	—	—	—	—	X
Isabel (United Kingdom)								
Isabel healthcare website – Isabel pro [131]	—	—	—	—	—	—	—	—
Bond et al [117]	—	X	—	—	—	—	—	—
Ramnarayan et al [162]	—	X	—	—	—	—	—	—
Ramnarayan et al [163]	—	X	—	—	—	—	—	—
Carlson et al [164]	—	X	—	—	—	—	—	—
Graber et al [165]	—	—	—	—	—	—	X	—
Graber et al [166]	—	X	—	—	—	—	—	—
Ramnarayan et al [167]	—	X	—	—	—	—	—	—
Bavdekar et al [168]	—	X	—	—	—	—	—	—
Ramnarayan et al [169]	—	X	—	—	—	—	—	—
Semigran et al [20]	—	X	—	—	—	—	—	—
Meyer et al [170]	—	X	—	—	—	—	—	—
Meditel (United States)								
No website identified	—	—	—	—	—	—	—	—
Berner et al [145]	X	X	X	—	—	—	—	—
Hammersley et al [149]	X	X	X	—	—	—	—	—
Waxman et al [171]	X	X	X	—	—	—	—	—
Wexler et al [123]	X	X	X	—	—	—	—	X
Medial Early sign (United States/Israel)								
Earlysign website [132]	—	—	—	—	—	—	—	—
Kinar et al [53][e]	—	—	—	—	—	—	—	—
Birks et al [52][e]	—	—	—	—	—	—	—	—
Hornbrook et al [34][e]	—	—	—	—	—	—	—	—
Goshen et al [172]	—		X	—	—	—	—	—
Zack et al [173]	—	X	—	—	—	—	—	—
Cahn et al [174]	—	X	—	—	—	—	—	—
Multilevel Diagnosis Decision Support System (Spain)								
No website identified	—	—	—	—	—	—	—	—
Rodriguez-Gonzalez et al [124]	X	X	—	—	—	—	—	X
Online webGP (United Kingdom; later became eConsult)								
Emis health online-triage website [175][f]	—	—	—	—	—	—	—	—
Hurleygroup website [176][g]	—	—	—	—	—	—	—	—

Technology identified (origin) websites and associated academic studies	Not AI[a]	Not cancer	Not primary care based	Not early detection or diagnosis	Early research	Not published	Not primary research	<50 cases or controls
Edwards et al [133]	X	X	—	X	—	—	—	—
Carter et al [177]	X	X	—	X	—	—	—	—
Cowie et al [178]	X	X	—	X	—	—	—	—
Pepid (United States)								
Pepid website [125][h]	N/C	—	—	—	—	—	—	—
Bond et al [117]	X	X	X	—	—	—	—	—
Problem Knowledge Couplers (PKC; United States)								
No website identified	—	—	—	—	—	—	—	—
Apkon et al [126]	X	—	—	X	—	—	—	—
Quick Medical Reference (QMR) (United States; developed from Internist-1)								
No website identified	—	—	—	—	—	—	—	—
Arene et al [179]	X	—	X	—	—	—	—	X
Bacchus et al [180]	X	—	X	—	—	—	—	X
Bankowitz et al [181]	X	—	X	—	—	—	—	X
Berner et al [145]	X	X	X	—	—	—	—	—
Berner et al [182]	X	—	—	—	—	—	—	X
Friedman et al [153]	X	—	X	—	—	—	—	X
Gozum et al [154]	X	—	X	—	—	—	—	X
Graber et al [155]	X	—	X	—	—	—	—	X
Miller et al [122]	X	—	X	—	—	—	—	X
Lemaire et al [183]	X	—	X	—	—	—	—	—
Reconsider (United States)								
No website identified	—	—	—	—	—	—	—	—
Nelson et al [127]	X	X	X	—	—	—	—	—
Symcat (United States)								
Symcat website [119]	—	—	—	—	—	X	—	—
Symptify (United States)								
Symptify website [134]	N/C	—	—	—	—	X	—	—
Symptomate (Poland)								
Symptomate website [135]	—	—	—	—	—	X	—	—
Unnamed								
No website identified	—	—	—	—	—	—	—	—
Liang H et al [136]	—	X	X	—	—	—	—	—

[a]AI: artificial intelligence.

[b]Not applicable or no data.

[c]Study excluded for the reason specified in the column label.

[d]N/C: not clear.

[e]These studies met the inclusion criteria of the systematic review and were therefore included.

[f]Edwards et al [133] suggests that this Egton Medical Information Systems (EMIS) application is powered by the eConsult system.

[g]Carter et al [177] suggests that this is the group who developed webGP.

[h]Several published studies are linked in the research section of the website, none involved use of the differential diagnosis or decision support tools. Some case studies audited the use of these tools.

Discussion

Principal Findings

We identified 16 studies reporting AI techniques that could facilitate the early detection of cancer and could be applied to the types of data found in primary care EHRs. However, heterogeneity of AI modalities, data set characteristics, outcome measures, conduct of these studies, and quality assessment meant that we were unable to draw strong conclusions about the utility of these techniques in primary care settings. There was a notable paucity of evidence on performance using primary care data. Coupled with the lack of evidence on implementation barriers or cost-effectiveness, this may help explain why AI techniques have not been adopted widely into primary care clinical practice to date. The study by Kinar et al [48] and its subsequent validation in independent data sets [34,52,53], including primary care data sets, is a valuable example of a staged evaluation of an AI technique from early development, via validation data sets, to evaluation in the population for intended use [22]. The work by Kop and collaborators [49,115,184] also represents a good example of the staged development of an AI technique, with sequential peer-reviewed, published evaluations at each stage.

We also identified 21 commercial AI technologies, many of which have not been evaluated and reported in peer-reviewed, published studies. Many other technologies that were patient-facing and designed for the triage of symptoms were identified but had not been applied to EHRs. Eight of these technologies appeared to be based on newer machine learning AI techniques, with the majority appearing to be driven by knowledge-based decision tree algorithms. Only one of the identified technologies has been evaluated specifically for cancer, although it may be more efficacious for these technologies to be very general in scope and to be widely used, rather than to have a narrow focus on cancer alone. With wider adoption, these technologies have a greater potential for raising patient and clinician awareness of cancer. However, it remains important to fully understand their diagnostic accuracy and safety, including for the triage of potential cancer symptoms. AI technologies applied to EHRs are potentially useful for primary care clinicians; however, they need to be designed in a way that is appropriate for the type and origin of the data found in primary care EHRs and to have been thoroughly and transparently evaluated in the population the technology is intended for.

Strengths and Limitations

The strengths of this systematic review include the following: a broad and inclusive search strategy to avoid missing studies; guidance of an international expert panel in the development of the protocol and search strategy; independent screening, quality assessment, and data extraction processes; followed PRISMA guidance; and a parallel scoping review for commercial AI technologies. As only a few heterogeneous studies were identified, it was not possible to synthesize the data and evaluate the utility of these AI techniques. Furthermore, only one commercially available AI technology was identified via the systematic review. Many of the technologies identified in the parallel scoping review lacked sufficient academic detailing and evidence for their accuracy or safety. This is a rapidly evolving research area, which will require further review over time.

Conclusions

Worldwide, there is a great deal of interest in AI techniques and their potential in medicine, not least in the United Kingdom where politicians and NHS leaders have publicly prioritized the incorporation of AI into clinical settings. Our findings support those of Kueper et al [17], namely, that although some AI techniques have good initial validation reports, they have not yet been through the steps for full application in clinical practice. Validation using independent data is preferable to splitting a single data set [185] and could be the next step in the development of many AI techniques identified in this review. Much of the research is at an early stage, with variable reporting and conduct, and requires further validation in prospective clinical settings and assessment of cost-effectiveness after clinical implementation before it can be incorporated into daily practice safely and effectively [186].

Consensus is required on how AI techniques designed for clinical use should be developed and validated to ensure their safety for patients and clinicians in their intended settings. Good internal and external validity is required in these experiments to avoid bias, most notably spectrum bias [187] and distributional shift [16], and to ensure that the appropriate data are used to develop the AI technique in keeping with its anticipated clinical setting and diagnostic function. The CanTest framework provides an outline for further studies aiming to develop this evidence base for AI techniques in clinical settings; to prove their safety and efficacy to commissioners, clinicians, and patients; and to enable them to be implemented in clinical practice [22]. Prospective evaluation in the clinical setting for which the AI technique is intended is essential: AI aimed at primary care clinics must be evaluated in primary care settings, where cancer prevalence is low compared with specialist settings, to accurately evaluate their future performance [187,188]. Further research around the acceptability of AI techniques for patients and clinicians and their cost-effectiveness will also be important to facilitate rapid implementation. Once these AI techniques are ready for implementation, they will require careful design to ensure effective integration into health information systems [189]. Data governance and protection must also be addressed, as they may present significant barriers to the implementation of these technologies [190,191].

In conclusion, AI techniques have the potential to aid the interpretation of patient-reported symptoms and clinical signs and to support clinical management, doctor-patient communication, and informed decision making. Ultimately, in the context of early cancer detection, these techniques may help reduce missed diagnostic opportunities and improve safety netting. However, although there are a few good examples of staged validation of these AI techniques, most of the research is at an early stage. We found numerous examples of the implementation of AI technologies without any or sufficient evidence for their accuracy or safety. Further research is required to build up the evidence base for AI techniques applied to EHRs

and to reassure commissioners, clinicians, and patients that they are safe and effective enough to be incorporated into routine clinical practice.

Acknowledgments

This research was funded by the National Institute for Health Research (NIHR) Policy Research Programme, conducted through the Policy Research Unit in Cancer Awareness, Screening, and Early Diagnosis, PR-PRU-1217-21601. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. This work was also supported by the CanTest Collaborative (funded by Cancer Research UK C8640/A23385), of which FW and WH are directors and JE, HS, and NdW are associate directors. HS is additionally supported by the Houston Veterans Administration Health Services Research and Development Center for Innovations in Quality, Effectiveness, and Safety (CIN13-413) and the Agency for Healthcare Research and Quality (R01HS27363). The funding sources had no role in the study design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit for publication. The authors would like to thank Isla Kuhn, Reader Services Librarian, University of Cambridge Medical Library, for her help in developing the search strategy.

Authors' Contributions

OJ developed the protocol, completed the search, screened the articles for inclusion, extracted the data, synthesized the findings, interpreted the results, and drafted the manuscript. NC screened the articles for inclusion, extracted the data, and critically revised the manuscript. SS screened the articles for inclusion, extracted the data, and critically revised the manuscript. WH developed the protocol, interpreted the results, and critically revised the manuscript. SD, JE, HS, and NdW critically revised the manuscript. FW developed the protocol, synthesized the findings, interpreted the results, and critically revised the manuscript. All authors approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Protocol for the study.
[DOCX File , 34 KB - jmir_v23i3e23483_app1.docx]

Multimedia Appendix 2
Search strategies.
[DOCX File , 16 KB - jmir_v23i3e23483_app2.docx]

Multimedia Appendix 3
Results of the full-text article review.
[DOCX File , 38 KB - jmir_v23i3e23483_app3.docx]

Multimedia Appendix 4
Supplementary information to table 1.
[DOCX File , 36 KB - jmir_v23i3e23483_app4.docx]

References

1. Cancer statistics for the UK. Cancer Research UK. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics-for-the-uk [accessed 2020-11-30]
2. Hamilton W. Diagnosing symptomatic cancer in the NHS. Br Med J 2015 Oct 13;351:5311. [doi: 10.1136/bmj.h5311] [Medline: 26466605]
3. Coleman M, Forman D, Bryant H, Butler J, Rachet B, Maringe C, et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. The Lancet 2011 Jan;377(9760):127-138. [doi: 10.1016/s0140-6736(10)62231-3]
4. Hiom SC. Diagnosing cancer earlier: reviewing the evidence for improving cancer survival. Br J Cancer 2015 Mar 31;112 Suppl 1(S1):S1-S5 [FREE Full text] [doi: 10.1038/bjc.2015.23] [Medline: 25734391]
5. Garbe C, Peris K, Hauschild A, Saiag P, Middleton M, Bastholt L, European Dermatology Forum (EDF), European Association of Dermato-Oncology (EADO), European Organisation for ResearchTreatment of Cancer (EORTC). Diagnosis and treatment of melanoma. European consensus-based interdisciplinary guideline - update 2016. Eur J Cancer 2016 Aug;63:201-217. [doi: 10.1016/j.ejca.2016.05.005] [Medline: 27367293]

6. Lyratzopoulos G, Wardle J, Rubin G. Rethinking diagnostic delay in cancer: how difficult is the diagnosis? Br Med J 2014 Dec 09;349(dec09 3):7400-7400 [FREE Full text] [doi: 10.1136/bmj.g7400] [Medline: 25491791]
7. Isabel Differential Diagnosis Generator. Isabel Healthcare. 2018. URL: https://www.isabelhealthcare.com [accessed 2020-11-30]
8. Artificial intelligence. Babylon Health. URL: https://www.babylonhealth.com/ai [accessed 2020-11-30]
9. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. J Am Med Assoc 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
10. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature 2020 Jan 1;577(7788):89-94. [doi: 10.1038/s41586-019-1799-6] [Medline: 31894144]
11. Li Z, Yu L, Wang X, Yu H, Gao Y, Ren Y, et al. Diagnostic performance of mammographic texture analysis in the differential diagnosis of benign and malignant breast tumors. Clin Breast Cancer 2018 Aug;18(4):621-627. [doi: 10.1016/j.clbc.2017.11.004] [Medline: 29199085]
12. Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. The Lancet 2018 Dec;392(10162):2388-2396. [doi: 10.1016/s0140-6736(18)31645-3]
13. Shafique S, Tehsin S. Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. Technol Cancer Res Treat 2018 Jan 01;17:1533033818802789 [FREE Full text] [doi: 10.1177/1533033818802789] [Medline: 30261827]
14. Preparing the healthcare workforce to deliver the digital future. The Topol Review.: NHS Health Education England; 2019. URL: https://topol.hee.nhs.uk/ [accessed 2020-11-30]
15. Artificial intelligence and primary care. Royal College of General Practitioners. URL: https://www.rcgp.org.uk/-/media/Files/CIRC/CIRC-AI-REPORT.ashx?la=en [accessed 2020-11-30]
16. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf 2019 Mar 12;28(3):231-237 [FREE Full text] [doi: 10.1136/bmjqs-2018-008370] [Medline: 30636200]
17. Kueper JK, Terry AL, Zwarenstein M, Lizotte DJ. Artificial intelligence and primary care research: a scoping review. Ann Fam Med 2020 May 01;18(3):250-258 [FREE Full text] [doi: 10.1370/afm.2518] [Medline: 32393561]
18. Millenson M, Baldwin J, Zipperer L, Singh H. Beyond Dr. Google: the evidence on consumer-facing digital tools for diagnosis. Diagnosis (Berl) 2018 Sep 25;5(3):95-105 [FREE Full text] [doi: 10.1515/dx-2018-0009] [Medline: 30032130]
19. Riches N, Panagioti M, Alam R, Cheraghi-Sohi S, Campbell S, Esmail A, et al. The Effectiveness of Electronic Differential Diagnoses (DDX) Generators: a systematic review and meta-analysis. PLoS One 2016 Mar 8;11(3):0148991 [FREE Full text] [doi: 10.1371/journal.pone.0148991] [Medline: 26954234]
20. Semigran H, Linder J, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. Br Med J 2015 Jul 08;351:3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
21. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 01;9(8):027743 [FREE Full text] [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
22. Walter FM, Thompson MJ, Wellwood I, Abel GA, Hamilton W, Johnson M, et al. Evaluating diagnostic strategies for early detection of cancer: the CanTest framework. BMC Cancer 2019 Jun 14;19(1):586 [FREE Full text] [doi: 10.1186/s12885-019-5746-6] [Medline: 31200676]
23. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4(1):1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
24. Jones O, Ranmuthu C, Prathivadi K, Saji S, Calanzani N, Emery J, et al. Establishing which modalities of artificial intelligence (AI) for the early detection and diagnosis of cancer are ready for implementation in primary care: a systematic review. Prospero: International prospective register of systematic reviews 2020 [FREE Full text] [doi: 10.15124/CRD42020176674]
25. McCarthy J, Minsky M, Rochester N, Shannon C. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. AI Magazine,27(4), 12. 2006. URL: https://doi.org/10.1609/aimag.v27i4.1904 [accessed 2021-01-25]
26. Muehlhauser L. What should we learn from past AI forecasts? Open Philanthropy Project. 2016. URL: https://www.openphilanthropy.org/focus/global-catastrophic-risks/potential-risks-advanced-artificial-intelligence/what-should-we-learn-past-ai-forecasts [accessed 2021-01-25]
27. AI in the UK?: Ready, Willing and Able. HOUSE OF LORDS: Select Committee on Artificial Intelligence. 2018. URL: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf [accessed 2021-01-25]
28. arXiv.org e-Print archive. Cornell University. URL: https://arxiv.org/ [accessed 2020-11-30]
29. Research. Google AI - research. URL: https://ai.google/research/ [accessed 2020-11-30]
30. Emerging technology, computer, and software research. Microsoft Research. URL: https://www.microsoft.com/en-us/research/ [accessed 2020-11-30]
31. Artificial intelligence. IBM Research. URL: https://www.research.ibm.com/artificial-intelligence/ [accessed 2020-11-30]

32. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536 [FREE Full text] [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
33. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews. ESRC Methods Program Swindon.: University of Lancaster; 2006. URL: https://tinyurl.com/19ok1m1j [accessed 2021-01-25]
34. Hornbrook MC, Goshen R, Choman E, O'Keeffe-Rosetti M, Kinar Y, Liles EG, et al. Early colorectal cancer detected by machine learning model using gender, age, and complete blood count data. Dig Dis Sci 2017 Oct 23;62(10):2719-2727. [doi: 10.1007/s10620-017-4722-8] [Medline: 28836087]
35. Survival three times higher when cancer is diagnosed early. Cancer Research UK. URL: https://www.cancerresearchuk.org/about-us/cancer-news/press-release/2015-08-10-survival-three-times-higher-when-cancer-is-diagnosed-early [accessed 2019-12-17]
36. NHS Long Term Plan. URL: https://www.longtermplan.nhs.uk/ [accessed 2021-02-08]
37. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Feb 02;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
38. Marchetti MA, Codella NC, Dusza SW, Gutman DA, Helba B, Kalloo A, International Skin Imaging Collaboration. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. J Am Acad Dermatol 2018 Feb;78(2):270-277.e1 [FREE Full text] [doi: 10.1016/j.jaad.2017.08.016] [Medline: 28969863]
39. ALzubi JA, Bharathikannan B, Tanwar S, Manikandan R, Khanna A, Thaventhiran C. Boosted neural network ensemble classification for lung cancer disease diagnosis. Applied Soft Computing 2019 Jul;80:579-591. [doi: 10.1016/j.asoc.2019.04.031]
40. Chang C, Hsu M. The study that applies artificial intelligence and logistic regression for assistance in differential diagnostic of pancreatic cancer. Expert Systems with Applications 2009 Sep;36(7):10663-10672. [doi: 10.1016/j.eswa.2009.02.046]
41. Cooper JA, Parsons N, Stinton C, Mathews C, Smith S, Halloran SP, et al. Risk-adjusted colorectal cancer screening using the FIT and routine screening data: development of a risk prediction model. Br J Cancer 2018 Jan 2;118(2):285-293 [FREE Full text] [doi: 10.1038/bjc.2017.375] [Medline: 29096402]
42. Cowley J. The use of knowledge discovery databases in the identification of patients with colorectal cancer. University of Hull [Dissertation]. 2012 Jul 01. URL: https://core.ac.uk/download/pdf/18526844.pdf [accessed 2020-11-30]
43. Daqqa K, Maghari A. Prediction and diagnosis of leukemia using classification algorithms. In: Proceedings of 8th International Conference on Information Technology. 2017 Presented at: 8th International Conference on Information Technology (ICIT); May 17-18, 2017; Amman, Jordan p. 638-643. [doi: 10.1109/icitech.2017.8079919]
44. Gorynski K, Safian I, Gradzki W. Artificial neural networks approach to early lung cancer detection. Cent Eur J Med 2014;9(5):632-641. [doi: 10.2478/s11536-013-0327-6]
45. Hart GR, Roffman DA, Decker R, Deng J. A multi-parameterized artificial neural network for lung cancer risk prediction. PLoS One 2018 Oct 24;13(10):0205264 [FREE Full text] [doi: 10.1371/journal.pone.0205264] [Medline: 30356283]
46. Kalra P, Togami J, Bansal BSG, Partin AW, Brawer MK, Babaian RJ, et al. A neurocomputational model for prostate carcinoma detection. Cancer 2003 Nov 01;98(9):1849-1854 [FREE Full text] [doi: 10.1002/cncr.11748] [Medline: 14584066]
47. Kang G, Ni Z. Research on early risk predictive model and discriminative feature selection of cancer based on real-world routine physical examination data. In: Proceedings of IEEE Int Conf Bioinforma Biomed BIBM. 2016 Presented at: IEEE Int Conf Bioinforma Biomed BIBM; 2016; Shenzhen, China p. 1512-1519. [doi: 10.1109/bibm.2016.7822746]
48. Kinar Y, Kalkstein N, Akiva P, Levin B, Half EE, Goldshtein I, et al. Development and validation of a predictive model for detection of colorectal cancer in primary care by analysis of complete blood counts: a binational retrospective study. J Am Med Inform Assoc 2016 Sep 15;23(5):879-890 [FREE Full text] [doi: 10.1093/jamia/ocv195] [Medline: 26911814]
49. Kop R, Hoogendoorn M, Teije AT, Büchner FL, Slottje P, Moons LM, et al. Predictive modeling of colorectal cancer using a dedicated pre-processing pipeline on routine electronic medical records. Comput Biol Med 2016 Sep 01;76:30-38. [doi: 10.1016/j.compbiomed.2016.06.019] [Medline: 27392227]
50. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016 May 17;6(1):26094 [FREE Full text] [doi: 10.1038/srep26094] [Medline: 27185194]
51. Payandeh M, Aeinfar M, Aeinfar V, Hayati M. A new method for diagnosis and predicting blood disorder and cancer using artificial intelligence (Artificial Neural Networks). Int J Hematol Stem Cell Res 2009;3(4):25-33. [doi: 10.1109/isie.2009.5213591]
52. Birks J, Bankhead C, Holt TA, Fuller A, Patnick J. Evaluation of a prediction model for colorectal cancer: retrospective analysis of 2.5 million patient records. Cancer Med 2017 Oct 21;6(10):2453-2460 [FREE Full text] [doi: 10.1002/cam4.1183] [Medline: 28941187]
53. Kinar Y, Akiva P, Choman E, Kariv R, Shalev V, Levin B, et al. Performance analysis of a machine learning flagging system used to identify a group of individuals at a high risk for colorectal cancer. PLoS One 2017 Feb 9;12(2):0171759 [FREE Full text] [doi: 10.1371/journal.pone.0171759] [Medline: 28182647]

54. Malik N, Idris W, Gunawan TS, Olanrewaju RF, Ibrahim SN. Classification of normal and crackles respiratory sounds into healthy and lung cancer groups. Int J Electr Comput Eng 2018 Jun 01;8(3):1530. [doi: 10.11591/ijece.v8i3.pp1530-1538]
55. Adams K, Sideris M, Papagrigoriadis S. Lunchtime Posters-Can we make “Straight to Test” decisions in Two Week Wait (2WW) patients with the help of an Artificial Neural Network (ANN)? Colorectal Dis 2014 Aug 22;16:41-68. [doi: 10.1111/codi.12643]
56. Ahmed A, Shah M, Wahid A, ul Islam S, Abbasi MK, Asghar MN. Big data analytics using neural networks for earlier cancer detection. J Med Imaging Hlth Inform 2017 Oct 01;7(6):1469-1474. [doi: 10.1166/jmihi.2017.2189]
57. Ahmed K, Emran AA, Jesmin T, Mukti RF, Rahman MZ, Ahmed F. Early detection of lung cancer risk using data mining. Asian Pac J Cancer Prev 2013;14(1):595-598 [FREE Full text] [doi: 10.7314/apjcp.2013.14.1.595] [Medline: 23534801]
58. Ahmen U, Rasool G, Zafar S, Maqbool HF. Fuzzy Rule Based Diagnostic System to Detect the Lung Cancer. 2018 Presented at: 2018 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube); November 12-13, 2018; Quetta, Pakistan URL: https://ieeexplore.ieee.org/document/8610976 [doi: 10.1109/ICECUBE.2018.8610976]
59. Alaa A, Moon K, Hsu W, van der Schaar M. ConfidentCare: a clinical decision support system for personalized breast cancer screening. IEEE Trans Multimedia 2016 Oct;18(10):1942-1955. [doi: 10.1109/TMM.2016.2589160]
60. Alharbi A, Tchier F, Rashidi M. Using a GeneticFuzzy algorithm as a computer aided breast cancer diagnostic tool. Asian Pac J Cancer Prev 2016;17(7):3651-3658 [FREE Full text] [Medline: 27510026]
61. Ayeldeen H, Elfattah MA, Shaker O, Hassanien AE, Kim TH. Case-Based Retrieval Approach of Clinical Breast Cancer Patients. 2015 Presented at: 2015 3rd International Conference on Computer, Information and Application; May 21-23, 2015; Yeosu, South Korea. [doi: 10.1109/CIA.2015.17]
62. Balachandran K. An efficient optimization based lung cancer pre-diagnosis system with aid of Feed Forward Back Propagation Neural Network (FFBNN). J Theor Appl Inf Technol 2013 Oct;56(2):263-271 [FREE Full text]
63. Bhar JA, George V, Malik B. Cloud Computing with Machine Learning Could Help Us in the Early Diagnosis of Breast Cancer. 2015 Presented at: Second International Conference on Advances in Computing and Communication Engineering; May 1-2, 2015; Dehradun, India. [doi: 10.1109/ICACCE.2015.62]
64. CHauhan R, Kaur H, Sharma S. A Feature Based Approach for Medical Databases. In: Proceedings of the International Conference on Advances in Information Communication Technology & Computing. 2016 Presented at: AICTC '16; Aug 2016; Bikaner, India. [doi: 10.1145/2979779.2979873]
65. Chen Y, Joo EM. Biomedical diagnosis and prediction using parsimonious fuzzy neural networks. 2012 Presented at: 38th Annual Conference on IEEE Industrial Electronics Society; December 24, 2012; Montreal, QC, Canada. [doi: 10.1109/IECON.2012.6388524]
66. Choudhury T, Kumar V, Nigam D, Vashisht V. Intelligent Classification of Lung & Oral Cancer through Diverse Data Mining Algorithms. Presented at: 2016 International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE); 2016; Ghaziabad p. 133-138. [doi: 10.1109/ICMETE.2016.24]
67. Çınar M, Engin M, Engin EZ, Ziya Ateşçi Y. Early prostate cancer diagnosis by using artificial neural networks and support vector machines. Expert Systems with Applications 2009 Apr;36(3):6357-6361. [doi: 10.1016/j.eswa.2008.08.010]
68. Del Grossi AA, De Mattos Senefonte HC, Quaglio VG. Prostate cancer biopsy recommendation through use of machine learning classification techniques. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Switzerland: Springer; 2014:710-721.
69. Durga S, Kasturi K. Lung disease prediction system using data mining techniques. J Adv Res in Dynamical and Contr Sys 2017;9(5):62-66 [FREE Full text]
70. Elhoseny M, Bian G, Lakshmanaprabu S, Shankar K, Singh AK, Wu W. Effective features to classify ovarian cancer data in internet of medical things. Computer Networks 2019 Aug;159:147-156. [doi: 10.1016/j.comnet.2019.04.016]
71. Elshazly HI, Elkorany AM, Hassanien AE. Ensemble-based classifiers for prostate cancer diagnosis. In: Proocedings of the 9th International Computer Engineering Conference: Today Information Society What’s Next?, ICENCO 2013.: IEEE Computer Society; 2013 Presented at: 9th International Computer Engineering Conference: Today Information Society What’s Next?, ICENCO 2013; 2013; 9th International Computer Engineering Conference: Today Information Society What’s Next?, ICENCO 2013 p. 49-54. [doi: 10.1109/ICENCO.2013.6736475]
72. Fan Y, Chaovalitwongse WA. Optimizing feature selection to improve medical diagnosis. Ann Oper Res 2009 Jan 6;174(1):169-183. [doi: 10.1007/s10479-008-0506-z]
73. Gaebel J, Cypko MA, Lemke HU. Accessing patient information for probabilistic patient models using existing standards. Stud Health Technol Inform 2016;223:107-112. [Medline: 27139392]
74. Gao Z, Gong J, Qin Q, Lin J. [Application of support vector machine in the detection of early cancer]. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi 2005 Oct;22(5):1045-1048. [Medline: 16294750]
75. Gelnarová E, Šafařík L. Comparison of three statistical classifiers on a prostate cancer data. Neural Network World 2005;15(4):311-318 [FREE Full text]
76. Ghaderzadeh M. Clinical decision support system for early detection of prostate cancer from benign hyperplasia of prostate. In: Proceedings of the 14th World Congress on Medical and Health Informatics, Pts 1 and 2. 2013 Presented at: Proceedings of the 14th World Congress on Medical and Health Informatics, Pts 1 and 2; 2013; Netherlands p. 928. [doi: 10.3233/978-1-61499-289-9-928]

77. Ghany KKA, Ayeldeen H, Zawbaa HM, Shaker O, IEEE. A rough set-based reasoner for medical diagnosis. In: Proceedings of the International Conference on Green Computing and Internet of Things. 2015 Presented at: International Conference on Green Computing and Internet of Things; 2015; Beni-Suef University, Egypt p. 429-434. [doi: 10.1109/ICGCIoT.2015.7380502]
78. Ghany KKA, Ayeldeen H, Zawbaa HM, Shaker O, Ayedeen G, IEEE. Diagnosis of breast cancer using secured classifiers. In: Proceedings of the International Conference on Electrical and Computing Technologies and Applications. 2017 Presented at: International Conference on Electrical and Computing Technologies and Applications; 2017; Beni-Suef University, Egypt p. 680-684. [doi: 10.1109/ICECTA.2017.8251947]
79. Goraneseu F, Gorunescu M, El-Darzi E, Ene M, Gorunescu S. Statistical comparison of a probabilistic neural network approach in hepatic cancer diagnosis. In: EUROCON 2005 - The International Conference on Computer as a Tool. 2005 Presented at: EUROCON 2005 - The International Conference on Computer as a Tool; 2005; Belgrade; Yugoslavia p. 237-240. [doi: 10.1109/EURCON.2005.1629904]
80. Gorunescu F, Belciug S. Boosting backpropagation algorithm by stimulus-sampling: application in computer-aided medical diagnosis. J Biomed Inform 2016 Oct;63:74-81 [FREE Full text] [doi: 10.1016/j.jbi.2016.08.004] [Medline: 27498068]
81. Gorunescu M, Gorunescu F, Revett K. A neural computing-based approach for the early detection of hepatocellular carcinoma. Proceedings of World Academy of Science, Engineering and Technology 2006;17:65 [FREE Full text]
82. Govinda K, Singla K, Jain K. Fuzzy based uncertainty modeling of Cancer Diagnosis System. In: Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS). 2017 Presented at: International Conference on Intelligent Sustainable Systems (ICISS); Dec 7-8, 2017; Palladam, India p. 740-743. [doi: 10.1109/ISS1.2017.8389272]
83. Halpern Y, Horng SK, Choi Y, Sontag D. Electronic medical record phenotyping using the anchor and learn framework. J Am Med Inform Assoc 2016 Jul;23(4):731-740 [FREE Full text] [doi: 10.1093/jamia/ocw011] [Medline: 27107443]
84. Hart GR, Roffman DA, Decker R. Scientific abstracts and sessions. Med Phys 2018 Jun 11;45(6):e120-e706. [doi: 10.1002/mp.12938]
85. Hornbrook MC, Goshen R, Choman E, O'Keeffe-Rosetti M, Kinar Y, Liles EG, et al. Correction to: early colorectal cancer detected by machine learning model using gender, age, and complete blood count data. Dig Dis Sci 2018 Jan;63(1):270. [doi: 10.1007/s10620-017-4859-5] [Medline: 29181742]
86. Hsu JL, Hung PC, Lin HY, Hsieh CH. Applying under-sampling techniques and cost-sensitive learning methods on risk assessment of breast cancer. J Med Syst 2015 Apr;39(4):210. [doi: 10.1007/s10916-015-0210-x] [Medline: 25712814]
87. Ilhan HO, Celik E. The mesothelioma disease diagnosis with artificial intelligence methods. In: Proceedings of the 10th International Conference on Application of Information and Communication Technologies (AICT). 2016 Presented at: 2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT); Oct 12-14, 2016; Baku, Azerbaijan. [doi: 10.1109/ICAICT.2016.7991825]
88. Ji Z, Wang B. Identifying potential clinical syndromes of hepatocellular carcinoma using PSO-based hierarchical feature selection algorithm. Biomed Res Int 2014;2014:127572 [FREE Full text] [doi: 10.1155/2014/127572] [Medline: 24745007]
89. Kong Q, Wang D, Wang Y, Jin Y, Jiang B. Multi-objective neural network-based diagnostic model of prostatic cancer. System Engineering Theory and Practice 2018;38(2):532-544. [doi: 10.12011/1000-6788(2018)02-0532-13]
90. Kou L, Yuan Y, Sun J, Lin Y. Prediction of cancer based on mobile cloud computing and SVM. In: Proceedings of the International Conference on Dependable Systems and Their Applications (DSA). 2017 Presented at: 2017 International Conference on Dependable Systems and Their Applications (DSA); 2017; Beijing, China. [doi: 10.1109/DSA.2017.20]
91. Kshivets O. P2.11-13 Precise early detection of lung cancer and blood cell circuit. J Thoracic Oncol 2018 Oct;13(10):S783. [doi: 10.1016/j.jtho.2018.08.1360]
92. Liu S, Gaudiot J, Cristini V. Prototyping virtual cancer therapist (VCT): a software engineering approach. Conf Proc IEEE Eng Med Biol Soc 2006;2006:5424-5427. [doi: 10.1109/IEMBS.2006.259230] [Medline: 17945900]
93. Liu Y, Pan Q, Zhou Z. Improved feature selection algorithm for prognosis prediction of primary liver cancer. In: Intelligence Science II. Switzerland: Springer; 2018:422-430.
94. Meng J, Zhang R, Chen D. Utilizing narrative text from electronic health records for early warning model of chronic disease. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2018 Presented at: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC); Oct 7-10, 2018; Miyazaki, Japan. [doi: 10.1109/SMC.2018.00713]
95. Mesrabadi HA, Faez K. Improving early prostate cancer diagnosis by using Artificial Neural Networks and Deep Learning. In: Proceedings of the 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS). 2018 Presented at: 2018 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS); Dec 25-27, 2018; Tehran, Iran. [doi: 10.1109/ICSPIS.2018.8700542]
96. Morgado P, Vicente H, Abelha A, Machado J, Neves J, Neves J. A case-based approach to colorectal cancer detection. In: Information Science and Applications 2017. Singapore: Springer; 2017:433-442.
97. Nalluri MR, Roy DS. Hybrid disease diagnosis using multiobjective optimization with evolutionary parameter optimization. J Healthc Eng 2017;2017:5907264 [FREE Full text] [doi: 10.1155/2017/5907264] [Medline: 29065626]
98. Nikitaev VG, Pronichev AN, Nagornov OV, Zaytsev SM, Polyakov EV, Romanov NA, et al. Decision support system in urologic cancer diagnosis. J Phys : Conf Ser 2019 Apr 16;1189:012032. [doi: 10.1088/1742-6596/1189/1/012032]

99. Polat K, Senturk U. A Novel ML Approach to Prediction of Breast Cancer: Combining of mad normalization, KMC based feature weighting and AdaBoostM1 classifierA novel ML approach to prediction of breast cancer: combining of mad normalization, KMC based feature weighting and AdaBoostM1 classifier. In: 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). 2018 Presented at: 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT); Oct 19-21, 2018; Ankara, Turkey. [doi: 10.1109/ISMSIT.2018.8567245]
100. Rahman A, Muniyandi RC. Feature selection from colon cancer dataset for cancer classification using Artificial Neural Network. Int J Adv Sci Engi and Info Tech. 2018. URL: https://www.researchgate.net/publication/328924307_Feature_selection_from_colon_cancer_dataset_for_cancer_classification_using_Artificial_Neural_Network [accessed 2021-02-08]
101. Ramya Devi M, Gomathy B. An intelligent system for the detection of breast cancer using feature selection and PCA methods. Int J Appl Engi Res. 2015. URL: https://www.researchgate.net/publication/283232820_An_intelligent_system_for_the_detection_of_breast_cancer_using_feature_selection_and_PCA_methods [accessed 2021-02-08]
102. Richter AN, Khoshgoftaar TM. Melanoma risk modeling from limited positive samples. Netw Model Anal Health Inform Bioinforma 2019 Apr 4;8(1):-. [doi: 10.1007/s13721-019-0186-4]
103. Safdari R, Arpanahi H, Langarizadeh M, Ghazisaiedi M, Dargahi H, Zendehdel K. Design a fuzzy rule-based expert system to aid earlier diagnosis of gastric cancer. Acta Inform Med 2018;26(1):19. [doi: 10.5455/aim.2018.26.19-23]
104. Shalev V, Kinar Y, Kalkstein N, Akiva P, Half E, Goldshtein I, et al. Computational analysis of blood counts significantly increases detection rate of gastric and colorectal cancers: PR0195 Esophageal, Gastric and Duodenal Disorders. J Gastroenterol and Hepatol 2013:761-762.
105. Sobar, Machmud R, Wijaya A. Behavior determinant based cervical cancer early detection with machine learning algorithm. Advanced Science Letters 2016;22(10):3120-3123. [doi: 10.1166/asl.2016.7980]
106. Soliman THA, Mohamed R, Sewissy AA. A hybrid analytical hierarchical process and deep neural networks approach for classifying breast cancer. In: Proceedings of the 11th International Conference on Computer Engineering & Systems (ICCES). 2016 Presented at: 2016 11th International Conference on Computer Engineering & Systems (ICCES); Dec 20-21, 2016; Cairo, Egypt. [doi: 10.1109/ICCES.2016.7822002]
107. Sushma Rani N, Srinivasa Rao P, Parimala P. An efficient statistical computation technique for health care big data using R. IOP Conf Ser : Mater Sci Eng 2017 Sep 07;225:012159. [doi: 10.1088/1757-899x/225/1/012159]
108. Wang D, Quek C, See Ng G. Ovarian cancer diagnosis using a hybrid intelligent system with simple yet convincing rules. Applied Soft Computing 2014 Jul;20:25-39. [doi: 10.1016/j.asoc.2013.12.018]
109. Wang G, Teoh JYC, Choi KS. Diagnosis of prostate cancer in a Chinese population by using machine learning methods. Annu Int Conf IEEE Eng Med Biol Soc 2018 Jul;2018:1-4. [doi: 10.1109/EMBC.2018.8513365] [Medline: 30440319]
110. Xu W, Zhang R, Qimin E, Liu J, Laing C. [The clinical application of data mining in laryngeal cancer]. Lin Chung Er Bi Yan Hou Tou Jing Wai Ke Za Zhi 2015 Jul;29(14):1272-1275. [Medline: 26672241]
111. Yasodha P, Ananthanarayanan NR. Analysing big data to build knowledge based system for early detection of ovarian cancer. Indian J Sci and Tech 2015;8(14). [doi: 10.17485/ijst/2015/v8i14/65745]
112. Zangooei MH, Habibi J, Alizadehsani R. Disease Diagnosis with a hybrid method SVR using NSGA-II. Neurocomputing 2014 Jul;136:14-29. [doi: 10.1016/j.neucom.2014.01.042]
113. Zhang L, Wang H, Liang J, Wang J. Decision support in cancer base on fuzzy adaptive PSO for feedforward neural network training. In: Proceedings of the International Symposium on Computer Science and Computational Technology. 2008 Presented at: 2008 International Symposium on Computer Science and Computational Technology; Dec 20-22, 2008; Shanghai, China. [doi: 10.1109/iscsct.2008.73]
114. Zhang Z, Zhang H, Bast Jr RC. An application of artificial neural networks in ovarian cancer early detection. In: Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. 2000 Presented at: IEEE-INNS-ENNS International Joint Conference on Neural Networks; July 27, 2000; Como, Italy URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-
115. Hoogendoorn M, Szolovits P, Moons LM, Numans ME. Utilizing uncoded consultation notes from electronic medical records for predictive modeling of colorectal cancer. Artif Intell Med 2016 May;69:53-61 [FREE Full text] [doi: 10.1016/j.artmed.2016.03.003] [Medline: 27085847]
116. Moss S, Mathews C, Day TJ, Smith S, Seaman HE, Snowball J, et al. Increased uptake and improved outcomes of bowel cancer screening with a faecal immunochemical test: results from a pilot study within the national screening programme in England. Gut 2017 Sep 07;66(9):1631-1644. [doi: 10.1136/gutjnl-2015-310691] [Medline: 27267903]
117. Bond WF, Schwartz LM, Weaver KR, Levick D, Giuliano M, Graber ML. Differential diagnosis generators: an evaluation of currently available computer programs. J Gen Intern Med 2012 Feb 26;27(2):213-219 [FREE Full text] [doi: 10.1007/s11606-011-1804-8] [Medline: 21789717]
118. DXplain. DXplain. URL: http://www.mghlcs.org/projects/dxplain [accessed 2020-11-30]
119. Symcat Symptom Checker. Symcat Symptom Checker. URL: http://www.symcat.com/ [accessed 2020-11-30]

120. Heckerling PS, Elstein AS, Terzian CG, Kushner MS. The effect of incomplete knowledge on the diagnoses of a computer consultant system. Med Inform (Lond) 1991 Jul 12;16(4):363-370. [doi: 10.3109/14639239109067658] [Medline: 1762472]
121. Miller RA, Pople HE, Myers JD. Internist-I, an experimental computer-based diagnostic consultant for general internal medicine. N Engl J Med 1982 Aug 19;307(8):468-476. [doi: 10.1056/nejm198208193070803]
122. Miller RA, McNeil MA, Challinor SM, Masarie FE, Myers JD. The Internist-1/quick medical reference project--status report. West J Med 1986 Dec;145(6):816-822 [FREE Full text] [Medline: 3544509]
123. Wexler JR, Swender PT, Tunnessen WW, Oski FA. Impact of a system of computer-assisted diagnosis. Initial evaluation of the hospitalized patient. Am J Dis Child 1975 Feb 01;129(2):203-205. [doi: 10.1001/archpedi.1975.02120390037008] [Medline: 1091140]
124. Rodríguez-González A, Torres-Niño J, Mayer MA, Alor-Hernandez G, Wilkinson MD. Analysis of a multilevel diagnosis decision support system and its implications: a case study. Comput Math Methods Med 2012 Sep;2012(9):367345-367333 [FREE Full text] [doi: 10.1155/2012/367345] [Medline: 23320043]
125. Clinical decision support. PEPID. URL: https://www.pepid.com/ [accessed 2020-11-30]
126. Apkon M, Mattera JA, Lin Z, Herrin J, Bradley EH, Carbone M, et al. A randomized outpatient trial of a decision-support information technology tool. Arch Intern Med 2005 Nov 14;165(20):2388-2394. [doi: 10.1001/archinte.165.20.2388] [Medline: 16287768]
127. Nelson SJ, Blois MS, Tuttle MS, Erlbaum M, Harrison P, Kim H, et al. Evaluating RECONSIDER. J Med Syst 1985 Dec;9(5-6):379-388. [doi: 10.1007/bf00992575]
128. Our Solutions!. Abtrace. URL: https://www.abtrace.co/solution/ [accessed 2020-11-30]
129. C the signs. C the Signs. URL: https://cthesigns.co.uk/ [accessed 2020-11-30]
130. DocResponse. DocResponse. URL: https://www.docresponse.com/ [accessed 2020-11-30]
131. Isabel Pro - the DDx Generator. Isabel Healthcare. URL: https://uk.isabelhealthcare.com/products/isabel-pro-ddx-generator [accessed 2020-11-30]
132. Medial EarlySign. Medial EarlySign. URL: https://earlysign.com/ [accessed 2020-11-30]
133. Edwards HB, Marques E, Hollingworth W, Horwood J, Farr M, Bernard E, et al. Use of a primary care online consultation system, by whom, when and why: evaluation of a pilot observational study in 36 general practices in South West England. BMJ Open 2017 Nov 22;7(11):e016901 [FREE Full text] [doi: 10.1136/bmjopen-2017-016901] [Medline: 29167106]
134. Simptify. Simptify.com. URL: https://symptify.com/ [accessed 2020-11-30]
135. Check your symptoms online. Symptomate. URL: https://symptomate.com/ [accessed 2020-11-30]
136. Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med 2019 Mar;25(3):433-438. [doi: 10.1038/s41591-018-0335-9] [Medline: 30742121]
137. Zhelezniak V, Busbridge D, Shen A, Smith S, Hammerla N. Decoding decoders: finding optimal representation spaces for unsupervised similarity tasks. ICLR 2018 Work Track. arXiv.org. Preprint posted online September 5, 2018. URL: https://arxiv.org/abs/1805.03435 [accessed 2020-11-30]
138. Douglas L, Zarov I, Gourgoulias K, Lucas C, Hart C, Baker A, et al. A universal marginalizer for amortized inference in generative models. NIPS 2017 Work Adv Approx Bayesian Inference. Preprint posted online November 2, 2017. URL: https://arxiv.org/abs/1711.00695 [accessed 2020-11-30]
139. Smith S, Turban D, Hamblin S, Hammerla N. Offline bilingual word vectors, orthogonal transformations and the inverted softmax. arXiv.org. Preprint posted online February 13, 2017. URL: https://arxiv.org/abs/1702.03859 [accessed 2020-11-30]
140. NHS 111 Powered by Babylon - outcomes evaluation. Babylon Health. 2017. URL: https://assets.babylonhealth.com/nhs/NHS-111-Evaluation-of-outcomes.pdf [accessed 2020-11-30]
141. Middleton K, Butt M, Hammerla N, Hamblin S, Mehta K, Parsa A. Sorting out symptoms: design and evaluation of the 'babylon check' automated triage system. arXiv.org. Preprint posted online June 7, 2016. URL: https://arxiv.org/abs/1606.02041 [accessed 2020-11-30]
142. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain: an evolving diagnostic decision-support system. J Am Med Assoc 1987 Jul 03;258(1):67. [doi: 10.1001/jama.1987.03400010071030]
143. Barnett GO, Famiglietti KT, Kim RJ, Hoffer EP, Feldman MJ. DXplain on the internet. Proc AMIA Symp 1998:607-611 [FREE Full text] [Medline: 9929291]
144. Bauer BA, Lee M, Bergstrom L, Wahner-Roedler DL, Bundrick J, Litin S, et al. Internal medicine resident satisfaction with a diagnostic decision support system (DXplain) introduced on a teaching hospital service. Proc AMIA Symp 2002:31-35 [FREE Full text] [Medline: 12463781]
145. Berner ES, Webster GD, Shugerman AA, Jackson JR, Algina J, Baker AL, et al. Performance of four computer-based diagnostic systems. N Engl J Med 1994 Jun 23;330(25):1792-1796. [doi: 10.1056/nejm199406233302506]
146. Elhanan G, Socratous SA, Cimino JJ. Integrating DXplain into a clinical information system using the World Wide Web. Proc AMIA Annu Fall Symp 1996:348-352 [FREE Full text] [Medline: 8947686]
147. Elkin PL, Liebow M, Bauer BA, Chaliki S, Wahner-Roedler D, Bundrick J, et al. The introduction of a diagnostic decision support system (DXplain™) into the workflow of a teaching hospital service can decrease the cost of service for diagnostically challenging Diagnostic Related Groups (DRGs). Int J Med Inform 2010 Nov;79(11):772-777 [FREE Full text] [doi: 10.1016/j.ijmedinf.2010.09.004] [Medline: 20951080]

148. Feldman MJ, Octo Barnett G. An approach to evaluating the accuracy of DXplain. Computer Methods and Programs in Biomedicine 1991 Aug;35(4):261-266. [doi: 10.1016/0169-2607(91)90004-d]
149. Hammersley J, Cooney K. Evaluating the utility of available differential diagnosis systems. Proc Annu Symp Comput Appl Med Care 1988 Nov 9:229-231 [FREE Full text]
150. Hoffer EP, Feldman MJ, Kim RJ, Famiglietti KT, Barnett GO. DXplain: patterns of use of a mature expert system. AMIA Annu Symp Proc 2005:321-325 [FREE Full text] [Medline: 16779054]
151. London S. DXplain: a web-based diagnostic decision support system for medical students. Medical Reference Services Quarterly 1998 May 07;17(2):17-28. [doi: 10.1300/j115v17n02_02]
152. Elstein A, Friedman C, Wolf F, Murphy G, Miller J, Fine P, et al. Effects of a decision support system on the diagnostic accuracy of users: a preliminary report. J Am Med Inform Assoc 1996;3(6):422-428 [FREE Full text] [doi: 10.1136/jamia.1996.97084515] [Medline: 8930858]
153. Friedman CP, Elstein AS, Wolf FM, Murphy GC, Franz TM, Heckerling PS, et al. Enhancement of clinicians' diagnostic reasoning by computer-based consultation: a multisite study of 2 systems. J Am Med Assoc 1999 Nov 17;282(19):1851-1856. [doi: 10.1001/jama.282.19.1851] [Medline: 10573277]
154. Gozum ME. Emulating cognitive diagnostic skills without clinical experience: a report of medical students using Quick Medical Reference and Iliad in the diagnosis of difficult clinical cases. Proc Annu Symp Comput Appl Med Care 1994:991 [FREE Full text] [Medline: 7950096]
155. Graber MA, VanScoy D. How well does decision support software perform in the emergency department? Emerg Med J 2003 Sep 01;20(5):426-428 [FREE Full text] [doi: 10.1136/emj.20.5.426] [Medline: 12954680]
156. Lange L, Haak S, Lincoln M. Use of Iliad to improve diagnostic performance of nurse practitioner students. J Nurs Educ 1997;36(1):36-45. [doi: 10.3928/0148-4834-19970101-09]
157. Lau L, Warner H, Poulsen A. Research review: a computer-based diagnostic model for individual case review. Top Health Inf Manage 1995 Feb;15(3):67-79. [Medline: 10140306]
158. Li YC, Haug PJ, Lincoln MJ, Turner CW, Pryor TA, Warner HH. Assessing the behavioral impact of a diagnostic decision support system. Proc Annu Symp Comput Appl Med Care 1995:805-809 [FREE Full text] [Medline: 8563402]
159. Lincoln MJ, Turner CW, Haug PJ, Warner HR, Williamson JW, Bouhaddou O, et al. Iliad training enhances medical students' diagnostic skills. J Med Syst 1991 Feb;15(1):93-110. [doi: 10.1007/bf00993883]
160. Murphy GC, Friedman CP, Elstein AS, Wolf FM, Miller T, Miller JG. The influence of a decision support system on the differential diagnosis of medical practitioners at three levels of training. Proc AMIA Annu Fall Symp 1996:219-223 [FREE Full text] [Medline: 8947660]
161. Wolf FM, Friedman CP, Elstein AS, Miller JG, Murphy GC, Heckerling P, et al. Changes in diagnostic decision-making after a computerized decision support consultation based on perceptions of need and helpfulness: a preliminary report. Proc AMIA Annu Fall Symp 1997:263-267 [FREE Full text] [Medline: 9357629]
162. Ramnarayan P, Roberts GC, Coren M, Nanduri V, Tomlinson A, Taylor PM, et al. Assessment of the potential impact of a reminder system on the reduction of diagnostic errors: a quasi-experimental study. BMC Med Inform Decis Mak 2006 Apr 28;6(1):22 [FREE Full text] [doi: 10.1186/1472-6947-6-22] [Medline: 16646956]
163. Ramnarayan P, Winrow A, Coren M, Nanduri V, Buchdahl R, Jacobs B, et al. Diagnostic omission errors in acute paediatric practice: impact of a reminder system on decision-making. BMC Med Inform Decis Mak 2006 Nov 06;6:37 [FREE Full text] [doi: 10.1186/1472-6947-6-37] [Medline: 17087835]
164. Carlson J, Abel M, Bridges D, Tomkowiak J. The impact of a diagnostic reminder system on student clinical reasoning during simulated case studies. Simulation in Healthcare: J Society Simul Healthcare 2011;6(1):11-17. [doi: 10.1097/sih.0b013e3181f24acd]
165. Graber ML. Taking steps towards a safer future: measures to promote timely and accurate medical diagnosis. Am J Med 2008 May;121(5 Suppl):S43-S46. [doi: 10.1016/j.amjmed.2008.02.006] [Medline: 18440355]
166. Graber ML, Tompkins D, Holland JJ. Resources medical students use to derive a differential diagnosis. Med Teach 2009 Jun 27;31(6):522-527. [doi: 10.1080/01421590802167436] [Medline: 19811168]
167. Ramnarayan P, Cronje N, Brown R, Negus R, Coode B, Moss P, et al. Validation of a diagnostic reminder system in emergency medicine: a multi-centre study. Emerg Med J 2007 Sep 01;24(9):619-624 [FREE Full text] [doi: 10.1136/emj.2006.044107] [Medline: 17711936]
168. Bavdekar S, Pawar M. Evaluation of an Internet-Delivered Pediatric Diagnosis Support System (ISABEL®) in a Tertiary Care Center in India. Indian Pediatr 2005;42(11):91. [Medline: 16340049]
169. Ramnarayan P, Britto J. Paediatric clinical decision support systems. Arch Dis Child 2002 Nov;87(5):361-362 [FREE Full text] [doi: 10.1136/adc.87.5.361] [Medline: 12390900]
170. Meyer AND, Giardina TD, Spitzmueller C, Shahid U, Scott TMT, Singh H. Patient perspectives on the usefulness of an artificial intelligence-assisted symptom checker: cross-sectional survey study. J Med Internet Res 2020 Jan 30;22(1):14679 [FREE Full text] [doi: 10.2196/14679] [Medline: 32012052]
171. Waxman HS, Worley WE. Computer-assisted adult medical diagnosis: subject review and evaluation of a new microcomputer-based system. Medicine (Baltimore) 1990 May;69(3):125-136. [Medline: 2189054]

172. Goshen R, Choman E, Ran A, Muller E, Kariv R, Chodick G, et al. Computer-assisted flagging of individuals at high risk of colorectal cancer in a large health maintenance organization using the colonflag test. JCO Clinical Cancer Informatics 2018 Dec(2):1-8. [doi: 10.1200/cci.17.00130]
173. Zack CJ, Senecal C, Kinar Y, Metzger Y, Bar-Sinai Y, Widmer RJ, et al. Leveraging machine learning techniques to forecast patient prognosis after percutaneous coronary intervention. JACC Cardiovasc Interv 2019 Jul 22;12(14):1304-1311 [FREE Full text] [doi: 10.1016/j.jcin.2019.02.035] [Medline: 31255564]
174. Cahn A, Shoshan A, Sagiv T, Yesharim R, Goshen R, Shalev V, et al. Prediction of progression from pre-diabetes to diabetes: development and validation of a machine learning model. Diabetes Metab Res Rev 2020 Feb 14;36(2):e3252. [doi: 10.1002/dmrr.3252] [Medline: 31943669]
175. EMIS Health - online triage. EMIS Health. URL: https://www.emishealth.com/products/partner-products/online-triage/ [accessed 2020-11-30]
176. Hurley Group. Hurley Group. URL: http://hurleygroup.co.uk/ [accessed 2020-11-30]
177. Carter M, Fletcher E, Sansom A, Warren FC, Campbell JL. Feasibility, acceptability and effectiveness of an online alternative to face-to-face consultation in general practice: a mixed-methods study of webGP in six Devon practices. BMJ Open 2018 Feb 15;8(2):018688 [FREE Full text] [doi: 10.1136/bmjopen-2017-018688] [Medline: 29449293]
178. Cowie J, Calveley E, Bowers G, Bowers J. Evaluation of a digital consultation and self-care advice tool in primary care: a multi-methods study. Int J Environ Res Public Health 2018 May 02;15(5):896 [FREE Full text] [doi: 10.3390/ijerph15050896] [Medline: 29724040]
179. Arene I, Ahmed W, Fox M, Barr CE, Fisher K. Evaluation of quick medical reference (QMR) as a teaching tool. MD Comput 1998;15(5):323-326. [Medline: 9753979]
180. Bacchus CM, Quinton C, O'Rourke K, Detsky AS. A randomized crossover trial of quick medical reference (QMR) as a teaching tool for medical interns. J Gen Intern Med 1994 Nov;9(11):616-621. [doi: 10.1007/bf02600304]
181. Bankowitz R, McNeil M, Challinor S, Parker R, Kapoor W, Miller R. A computer-assisted medical diagnostic consultation service. Implementation and prospective evaluation of a prototype. Ann Intern Med 1989 May 15;110(10):824-832 [FREE Full text] [doi: 10.7326/0003-4819-110-10-824] [Medline: 2653156]
182. Berner ES, Maisiak RS, Cobbs CG, Taunton OD. Effects of a decision support system on physicians' diagnostic performance. J Am Med Inform Assoc 1999 Sep 01;6(5):420-427 [FREE Full text] [doi: 10.1136/jamia.1999.0060420] [Medline: 10495101]
183. Lemaire JB, Schaefer JP, Martin LA, Faris P, Ainslie MD, Hull RD. Effectiveness of the Quick Medical Reference as a diagnostic tool. Can Med Asso J 1999 Sep 21;161(6):725-728 [FREE Full text] [Medline: 10513280]
184. Kop R, Hoogendoorn M, Moons L, Numans M, ten Teije A. On the advantage of using dedicated data mining techniques to predict colorectal cancer. In: Holmes J, Bellazzi R, Sacchi L, Peek N, editors. Artificial Intelligence in Medicine. AIME 2015. Lecture Notes in Computer Science, vol 9105. Switzerland: Springer International Publishing; 2015:133-142.
185. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. Br Med J 2009 May 28;338(may28 1):605-605. [doi: 10.1136/bmj.b605] [Medline: 19477892]
186. Singh H, Sittig DF. A sociotechnical framework for Safety-Related Electronic Health Record Research Reporting: The SAFER Reporting framework. Annals of Internal Medicine 2020 Jun 02;172(11_Supplement):92-100. [doi: 10.7326/m19-0879]
187. Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. Br Med J 2016 Jun 22;353:3139 [FREE Full text] [doi: 10.1136/bmj.i3139] [Medline: 27334281]
188. Kanagasingam Y, Xiao D, Vignarajan J, Preetham A, Tay-Kearney M, Mehrotra A. Evaluation of artificial intelligence-based grading of diabetic retinopathy in primary care. JAMA Netw Open 2018 Sep 07;1(5):182665 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.2665] [Medline: 30646178]
189. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019 Oct 29;17(1):195 [FREE Full text] [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]
190. Forcier M, Gallois H, Mullan S, Joly Y. Integrating artificial intelligence into health care through data access: can the GDPR act as a beacon for policymakers? J Law Biosci 2019 Oct;6(1):317-335 [FREE Full text] [doi: 10.1093/jlb/lsz013] [Medline: 31666972]
191. European Parlimentary Research Service (EPRS). The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. STUDY: Panel for the Future of Science and Technology. 2020. URL: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf [accessed 2020-11-30]

Abbreviations

AI: artificial intelligence
AUROC: area under the receiver operating characteristic
CT: computed tomography
EHR: electronic health record
NHS: National Health Service

NIHR: National Institute for Health Research
NPV: negative predictive value
PPV: positive predictive value
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analysis
QUADAS-2: quality assessment of diagnostic accuracy studies-2

Edited by G Eysenbach; submitted 13.08.20; peer-reviewed by Y Liang, Y Chu, R Verheij; comments to author 01.10.20; revised version received 05.11.20; accepted 30.11.20; published 03.03.21.

Please cite as:
Jones OT, Calanzani N, Saji S, Duffy SW, Emery J, Hamilton W, Singh H, de Wit NJ, Walter FM
Artificial Intelligence Techniques That May Be Applied to Primary Care Data to Facilitate Earlier Diagnosis of Cancer: Systematic Review
J Med Internet Res 2021;23(3):e23483
URL: https://www.jmir.org/2021/3/e23483
doi:10.2196/23483
PMID:33656443

Original Paper

Future Medical Artificial Intelligence Application Requirements and Expectations of Physicians in German University Hospitals: Web-Based Survey

Oliver Maassen[1,2], MSc; Sebastian Fritsch[1,2,3], MD; Julia Palm[2,4], MSc; Saskia Deffge[1,2], MSc; Julian Kunze[1,2], MD; Gernot Marx[1,2], MD, Prof Dr, FRCA; Morris Riedel[2,3,5], Prof Dr; Andreas Schuppert[2,6], Prof Dr; Johannes Bickenbach[1,2], MD, Prof Dr

[1]Department of Intensive Care Medicine, University Hospital RWTH Aachen, Aachen, Germany

[2]SMITH Consortium of the German Medical Informatics Initiative, Leipzig, Germany

[3]Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich, Germany

[4]Institute of Medical Statistics, Computer and Data Sciences, Jena University Hospital, Jena, Germany

[5]School of Natural Sciences and Engineering, University of Iceland, Reykjavik, Iceland

[6]Institute for Computational Biomedicine II, University Hospital RWTH Aachen, Aachen, Germany

Corresponding Author:
Oliver Maassen, MSc
Department of Intensive Care Medicine
University Hospital RWTH Aachen
Pauwelsstraße 30
Aachen, 52074
Germany
Phone: 49 2418080444
Email: oliver.maassen@rwth-aachen.de

Abstract

Background: The increasing development of artificial intelligence (AI) systems in medicine driven by researchers and entrepreneurs goes along with enormous expectations for medical care advancement. AI might change the clinical practice of physicians from almost all medical disciplines and in most areas of health care. While expectations for AI in medicine are high, practical implementations of AI for clinical practice are still scarce in Germany. Moreover, physicians' requirements and expectations of AI in medicine and their opinion on the usage of anonymized patient data for clinical and biomedical research have not been investigated widely in German university hospitals.

Objective: This study aimed to evaluate physicians' requirements and expectations of AI in medicine and their opinion on the secondary usage of patient data for (bio)medical research (eg, for the development of machine learning algorithms) in university hospitals in Germany.

Methods: A web-based survey was conducted addressing physicians of all medical disciplines in 8 German university hospitals. Answers were given using Likert scales and general demographic responses. Physicians were asked to participate locally via email in the respective hospitals.

Results: The online survey was completed by 303 physicians (female: 121/303, 39.9%; male: 173/303, 57.1%; no response: 9/303, 3.0%) from a wide range of medical disciplines and work experience levels. Most respondents either had a positive (130/303, 42.9%) or a very positive attitude (82/303, 27.1%) towards AI in medicine. There was a significant association between the personal rating of AI in medicine and the self-reported technical affinity level (H_4=48.3, P<.001). A vast majority of physicians expected the future of medicine to be a mix of human and artificial intelligence (273/303, 90.1%) but also requested a scientific evaluation before the routine implementation of AI-based systems (276/303, 91.1%). Physicians were most optimistic that AI applications would identify drug interactions (280/303, 92.4%) to improve patient care substantially but were quite reserved regarding AI-supported diagnosis of psychiatric diseases (62/303, 20.5%). Of the respondents, 82.5% (250/303) agreed that there should be open access to anonymized patient databases for medical and biomedical research.

Conclusions: Physicians in stationary patient care in German university hospitals show a generally positive attitude towards using most AI applications in medicine. Along with this optimism comes several expectations and hopes that AI will assist

physicians in clinical decision making. Especially in fields of medicine where huge amounts of data are processed (eg, imaging procedures in radiology and pathology) or data are collected continuously (eg, cardiology and intensive care medicine), physicians' expectations of AI to substantially improve future patient care are high. In the study, the greatest potential was seen in the application of AI for the identification of drug interactions, assumedly due to the rising complexity of drug administration to polymorbid, polypharmacy patients. However, for the practical usage of AI in health care, regulatory and organizational challenges still have to be mastered.

(J Med Internet Res 2021;23(3):e26646) doi:10.2196/26646

KEYWORDS

artificial intelligence; AI; machine learning; algorithms; clinical decision support; physician; requirement; expectation; hospital care

Introduction

While a balance between the increasing amount of documented data on the one hand and the demographic change and aging populations on the other hand challenges our health care systems, big data and artificial intelligence (AI) in medicine offer a huge potential to relieve physicians from the increasing complexity of today's health care and information overload when treating patients [1,2]. Over the last decade, research on AI in medicine and biomedicine and the number of publications in these fields have substantially increased [3,4]. Research has come up with promising AI developments in general machine learning (ML) algorithms, for manifold applications to predict clinical events, to improve diagnoses accuracy as well as treatments, and to reduce the burden of disease [5,6]. Well-known examples for AI in medicine are the application of deep learning as a subfield of ML in medical imaging for disease detection from x-rays [7]. Thus, it is expected that AI in medical practice will meet higher expectations of medical treatment and physicians and will increase the efficiency of clinical care. AI is perceived as the next big thing that will sustainably change medicine towards precision and personalized medicine [8] and change health care and with it, the role of physicians. It is believed that physicians will not be replaced by AI, but AI will make lives easier and focussed where human interaction is really required.

Big Data and AI in Medicine: Definition and Application Areas

An ever-increasing amount of medical data is being recorded by monitoring patient care devices, enabling big data analysis in health care [9]. This paves the way for the application of different ML techniques like deep learning [7], traditional shallow neural networks, support vector machines, and random forests, which are specific models for using AI in practice [10]. Besides conventional ML techniques, deep learning in particular offers advantages for understanding and manipulating the highly relevant class of data, especially for images, language, and speech recognition [11]. However, while deep learning is superior for specific applications, it has limitations for other applications where conventional ML techniques are superior [12]. Examples include cases where large datasets are not available to study a specific medical condition, as deep learning generally requires a large dataset to perform well in practice.

Despite its widespread use, a holistic definition of the term "artificial intelligence" is challenging. This can be partly explained by the fact that it is a "high-level" term, often not mentioning concrete ML algorithms or models in a clear context. Examples of AI in medicine are AI applications to support diagnostic procedures, predict the course of the disease [13-17], enhance the potential of clinical decision support [18], and support the management of hospital workflows [19,20]. Thereby, AI offers the possibility to support physicians in delivering high-quality medicine and increasing medical care efficiency.

Preconditions for AI Development for Health Care

One essential precondition for the development of AI in medical practice is data availability to develop and train the algorithms. Therefore, the creation of research databases with consolidated anonymized patient data, ideally from multiple locations making clinical routine data available for so called "secondary usage," is desirable. One very prominent example of such a database is the freely accessible critical care database MIMIC-III (Medical Information Mart for Intensive Care) [21]. After accepting a data use agreement, researchers are granted unrestricted access for analysis. Since its publication in August 2015, the respective publication was cited more than 850 times, showing the great interest in this database and its wide scientific usage [22]. Nevertheless, due to different health care systems and information systems used for patient care, the structure and content of patient data from German hospitals are not in accordance with data from US hospitals. Consequently, there is an urgent need to establish research databases that are applicable to the situation in Germany.

Challenges of Big Data and AI in Medicine

Today, the widespread practical implementation of AI and AI-based decision support into hospital care has not yet become reality [23]. Especially in radiology and other medical imaging procedures, vendors of medical technology have predominantly integrated some kind of AI (eg, ML algorithms) into their products [24]. In addition, in the field of medical prediction, there are cases that have already been applied to electronic health records to complete prospective verification studies [25,26]. Cabitza et al [27] described the sociotechnical elements that must be mastered to successfully implement potentially effective AI in real-world clinical settings, which they call the "last mile" gap of AI implementation. Further multifaceted technical, regulatory, social, and human factors hinder the practical application of AI in clinical care [23].

Tremendous efforts have already been made to evaluate AI in health care and AI-enabled clinical decision support [28]. However, only a few medical AI applications are really used in clinical practice. As physicians are supposed to be the primary users of AI in medicine, there is a need to investigate physicians' requirements and expectations for future developments of AI in health care. In addition, regulatory questions, like the liability for medical errors, must be regulated if AI is to be used in physician practice and patient care [29].

Another often observed challenge in applying AI is the availability of sensitive patient datasets due to General Data Protection Regulation constraints, mostly when AI is used beyond just one health care organization. In this context, federated ML [30] is a promising approach to obtain powerful, accurate, safe, robust, and unbiased models by enabling multiple organizations to train collaboratively without the need to exchange or centralize datasets. Also, a federated ML approach can be useful where datasets within one organization might not be enough to train good models (eg, rare diseases). In this context, transfer learning approaches [31] are feasible, too, whereby medical AI models are created using a pretrained, state-of-the-art AI model from a different larger medical dataset or otherwise openly available data (eg, ImageNet dataset).

Both are not often applied in Germany today. To address the multifaceted complexity of AI, the purpose of our study was to evaluate the general perception towards AI in medicine among physicians, but also towards concrete application and the opinion on the usage of anonymized patient data for (bio)medical research and AI development.

To achieve this, we developed a web-based survey for physicians in German academic hospitals. The results should also help researchers and data scientists to better understand physicians' needs regarding AI systems and to boost their use in clinical care.

Methods

Study Design, Data Collection, and Recruitment

For the survey's conceptualization, open and explorative interviews were carried out with 3 junior and 3 senior physicians. The results were structured and then utilized to frame the survey questions in German. These questions were integrated into an open web-based survey (LimeSurvey) to be conducted among our study population in 8 German university hospitals consisting exclusively of physicians from the full range of medical disciplines. The local ethics committee and local data protection officer did not express objections to the operationalization of the web-based survey. On the first page of the survey, we informed the participants about the length, purpose, and expected time to fill in the questionnaire. As participants were free to participate and contribute to the study, we regarded the survey's completion as consent for the usage, analysis, and publication of the collected survey data. For verification and functionality validation, we performed a test phase of the online survey with 25 anesthesiologists and critical care physicians in June 2019. Minor adaptations were added in the final survey version before its link was sent to physicians in 8 university hospitals via email by local persons in charge. The survey was available for 19 weeks from June 2019 till October 2019. Within the data collection period, no content modifications nor bug fixes were necessary, and we did not identify any unforeseen events like system errors or server downtime.

The survey was separated into 2 sections. The first section was comprised of questions about AI in medicine, and the second contained general biographical questions.

A translated, English version of the survey is attached in Multimedia Appendix 1. We only included completely filled out questionnaires in the statistical analysis. However, this might include some questions that were not answered (no response).

The questions about AI in medicine are separated into 3 sections: (1) personal opinion about AI in health care (Q1.1-Q1.16; Q4), (2) fields of application of AI in medicine (Q2.1-Q2.25), (3) usage of anonymized patient data for research purposes (Q3.1-Q3.4)

The question groups 1, 2, and 3 were phrased as single-choice questions asking physicians about their personal view on given statements using a 4-point Likert scale without a neutral option. The first set of questions (Q1.1–Q1.16) in the survey explored the attitudes towards AI in medicine. The second set of questions focused on fields of AI application in medicine. In total, 25 AI applications were given, and physicians were asked to rate if the proposed applications could substantially improve patient care in the future. The third set of questions explored physicians' opinions on the secondary usage of anonymized patient data for research purposes (eg, for AI development for medical practice). Physicians were asked whether they agreed or disagreed with the statements on the usage of anonymized patient data for clinical and biomedical research. Finally, we asked physicians how positively or negatively they evaluated the use of AI in medicine on a 5-point Likert scale. This question (Q4) was assigned to the first section, "Personal opinion about AI in health care." Biographical answers were mostly conceptualized as closed-ended, single-choice questions (eg, demographic questions), but were also presented as multiple choice questions (eg, medical discipline and predominant workplace).

Statistical Analysis

For the statistical analysis, we stratified the potential fields of AI applications in medicine into 6 categories: (1) imaging procedures, (2) other diagnostic procedures, (3) intensive care unit (ICU)/anesthesia, (4) medication and therapy, (5) workflow support and education, (6) prognosis assessment (Textbox 1). Here, we partly reference peer-reviewed publications of AI algorithms, comparing AI algorithms with physicians cited by Topol [32]. In addition, we categorized the 33 medical disciplines into 6 categories for further subgroup analysis (see Table S1 in Multimedia Appendix 2). We analyzed most of the data descriptively using graphics produced by the R packages sjPlot [33] and ggplot2 [34]. Where appropriate in the survey data analysis, we conducted Kruskal-Wallis tests to investigate the relationship between AI rating and biographical data. All analyses were conducted in R version 4.0.3 [35].

Textbox 1. Categories of artificial intelligence (AI) applications in stationary hospital care.

- 1. Imaging procedures
  - 1.1 Analysis of x-rays, computed tomography (CT), magnetic resonance tomography (MRT), sonographies [36-38]
  - 1.2 Analysis of histopathologic fine cuts [39,40]
  - 1.3 Analysis of endoscopic pictures or videos [41-43]
  - 1.4 Analysis of dermatologic reflected light microscopy [44,45]
- 2. Other diagnostic procedures
  - 2.1 Analysis of electroencephalography (EEG)/electrocardiography (ECG) [46,47]
  - 2.2 Diagnosing rare diseases [48,49]
  - 2.3 Triage in emergency care [50,51]
  - 2.4 Diagnosing psychiatric diseases [52,53]
  - 2.5 Subspecification of hematologic diseases [54-56]
- 3. Intensive care unit (ICU)/anesthesia
  - 3.1 Early alarm of the deterioration of patient status [57]
  - 3.2 Reduction of false alarms in intensive care medicine [58]
  - 3.3 Automatic mechanical ventilation [59,60]
  - 3.4 Support of parenteral or enteral nutrition [61]
  - 3.5 Automatic anesthesia administration [62]
- 4. Medication and therapy
  - 4.1 Oncologic therapy planning [18,63]
  - 4.2 Antibiotic stewardship [64]
  - 4.3 Identification of drug interactions [65-67]
  - 4.4 Medication for geriatric patients [68]
  - 4.5 Medication for pediatric patients [69]
- 5. Workflow support and education
  - 5.1 Education and training of medical students and physicians [70]
  - 5.2 Workflow support in stationary hospital care [19,20]
  - 5.3 Medical recording or discharge letters [71,72]
- 6. Prognosis assessment
  - 6.1 Prediction of effects of therapeutic interventions [73]
  - 6.2 Assessment of prognosis of malignant diseases [13-15]
  - 6.3 Assessment of prognosis of nonmalignant diseases [16,17]

Results

Demographic and Professional Characteristics

The online survey was finished by 121 (121/303, 39.9%) female and 173 (173/303, 57.1%) male physicians (no response: 9/303, 3.0%). Their mean length of clinical work experience was 12.7 years. In particular, physicians from the age groups of 25-34 years (98/303, 32.3%), 35-44 years (103/303, 34.0%), and 45-54 years (69/303, 22.8%) participated. Physicians from a wide range of medical disciplines and from all proposed clinical hierarchical levels took part in the survey (Table 1). Additionally, all proposed operational areas (hospital ward, operating theater, outpatient clinic, ICU, office, laboratory, functional area, others areas) are represented in the survey (Table 1).

Table 1. Demographic and professional characteristics.

Characteristic	Values (n=303)
Age range (years), n (%)	
18-24	1 (0.3)
25-34	98 (32.3)
35-44	103 (34.0)
45-54	69 (22.8)
55-65	21 (6.9)
>65	3 (1.0)
No response	8 (2.6)
Gender, n (%)	
Female	121 (39.9)
Male	173 (57.1)
No response	9 (3.0)
Current occupation, n (%)	
Assistant physician	101 (33.3)
Medical specialist	49 (16.2)
Senior physician	108 (35.6)
Clinic director	28 (9.2)
Others	6 (2.0)
No response	11 (3.6)
Medical field or discipline, n (%)	
Anesthesiology/intensive care medicine	75 (24.8)
Internal medicine	53 (17.5)
Pediatrics	25 (8.3)
Surgery	22 (7.3)
Neurology	14 (4.6)
Dermatology	12 (4.0)
Microbiology, virology, infectiology	10 (3.3)
Psychiatry and psychotherapy	10 (3.3)
Psychosomatic medicine and psychotherapy	8 (2.6)
Neurosurgery	8 (2.6)
Ophthalmology	7 (2.3)
Pathology	7 (2.3)
Otorhinolaryngology	5 (1.7)
Child and adolescent psychiatry and psychotherapy	5 (1.7)
Laboratory medicine	5 (1.7)
Radiology	5 (1.7)
Urology	5 (1.7)
Other disciplines/ specialization	43 (14.2)
Predominant workplace, n (%)	
Hospital ward	123 (40.6)
Operating theater	106 (35.0)
Outpatient clinic	100 (33.0)

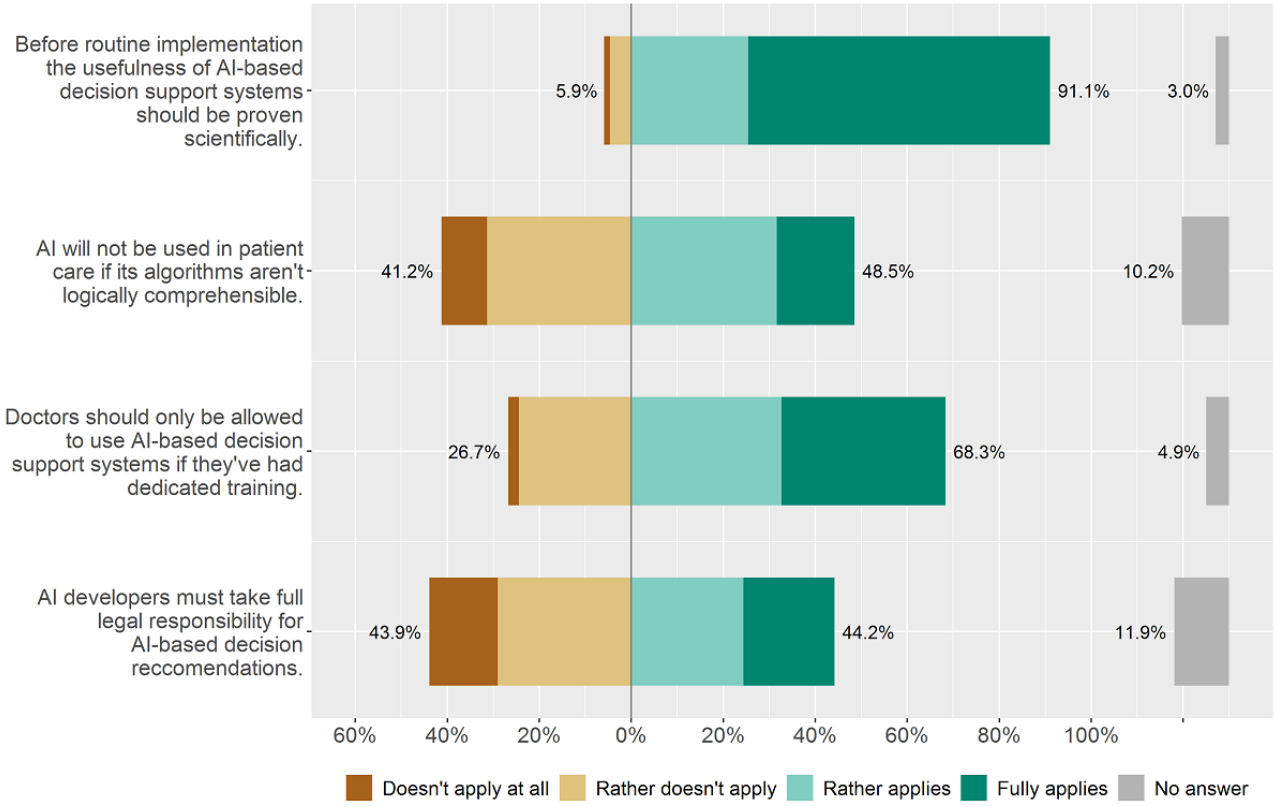
Characteristic	Values (n=303)
Intensive care unit	89 (29.4)
Office	50 (16.5)
Laboratory	33 (10.9)
Functional area	30 (9.9)
Others	15 (5.0)
Clinical professional experience (years), mean (SD)	12.7 (9.3)

Physicians' Attitudes Towards AI in Medicine

A majority of physicians reported either a positive (130/303, 42.9%) or a very positive attitude (82/303, 27.1%) towards AI in medicine (Q4; see Multimedia Appendix 3), representing more than two-thirds of the respondents; 18.2% (55/303) rated it neutral, and just 5.6% (17/303) rated it either negative or very negative.

As described in the Methods, we categorized the first question group into 3 subcategories. The first category focused on the rules and regulatory requirements of AI in medicine (see Figure 1). We found strong agreement (ie, "rather applies" and "fully applies") among physicians (276/303, 91.1%) for a scientific evaluation before the implementation of an AI-based system. Furthermore, the requirement for special "AI training" for physicians before usage of an AI-based decision support system was clearly favored (207/303, 68.3%).

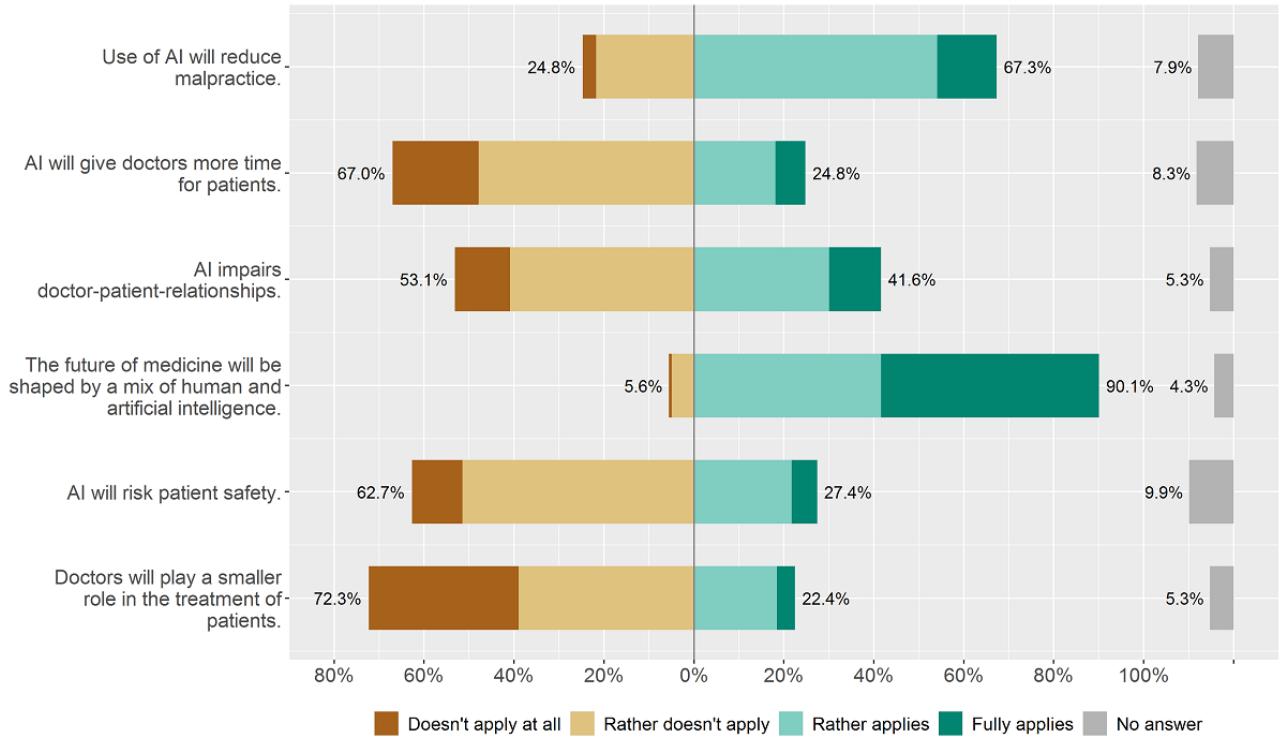
Figure 1. Rules and regulatory requirements of artificial intelligence (AI) usage in medicine.



The statements regarding the responsibility for AI decisions and the influence of algorithms' lack of logical comprehensibility returned a much less clear picture with answers split nearly in half.

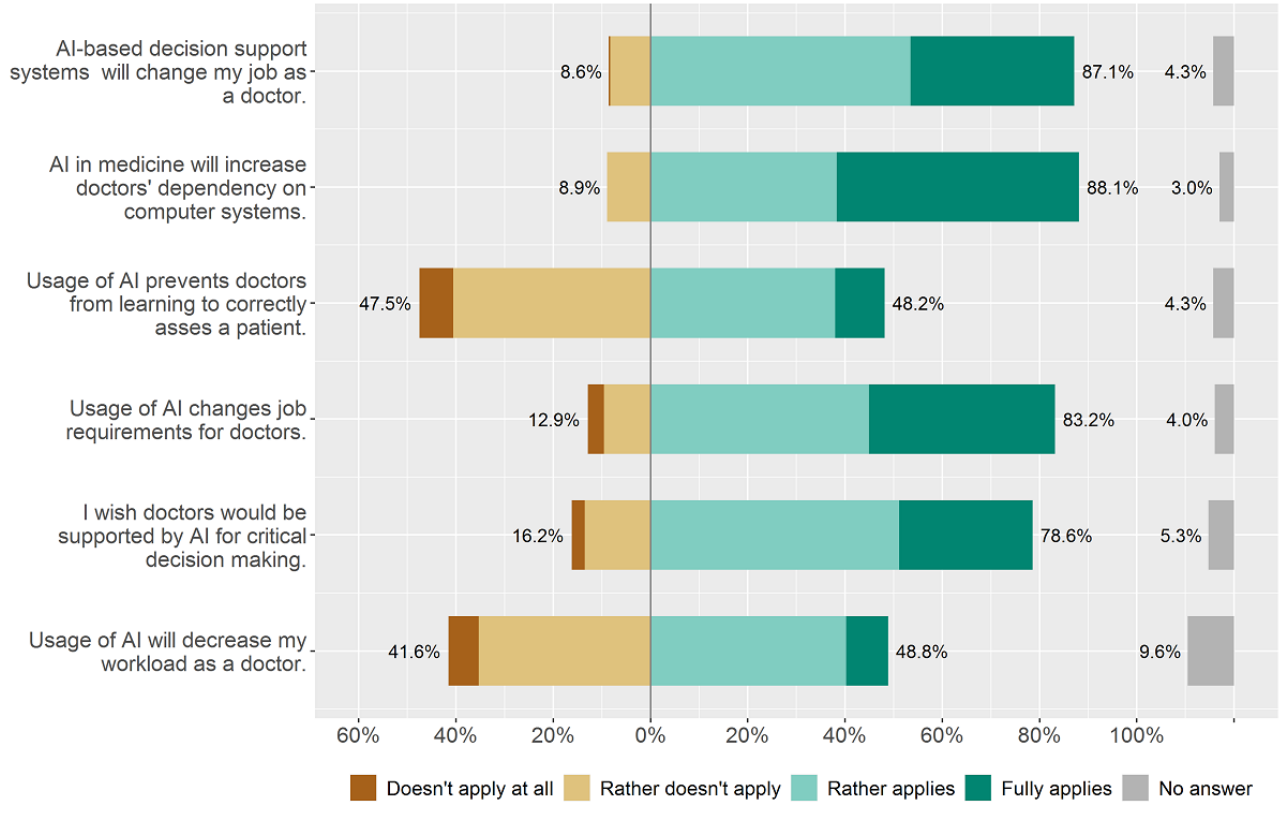
In the second subcategory of questions, "effect of AI on medical treatment" (see Figure 2), respondents mostly agreed with the statement "The future or medicine will be shaped by a mix of human and artificial intelligence" (273/303, 90.1%). Most physicians also expected a reduction of malpractice through the use of AI (204/303, 67.3%). At the same time, a majority didn't expect (ie, "doesn't apply at all" and "rather doesn't apply") that AI would give them more time for their patients (203/303, 67.0%) or that they would play a minor role in the treatment of patients (219/303, 72.3%).

Figure 2. Effect of artificial intelligence (AI) on medical treatment.



The third subcategory of questions (see Figure 3) focussed on the effect of AI in physicians' work. Here, respondents agreed that the usage of AI in health care would increase physicians' dependence on computer systems (267/303, 88.1%). They also agreed that AI-based decision support systems would change their work as a physician (264/303, 87.1%). The clear majority also anticipated a change of physicians' job requirements (252/303, 83.2%). For the statement "Usage of AI prevents doctors from learning to correctly assess a patient," agreement and disagreement were roughly evenly distributed (agreement: 146/303, 48.2%; disagreement: 144/303, 47.5%).

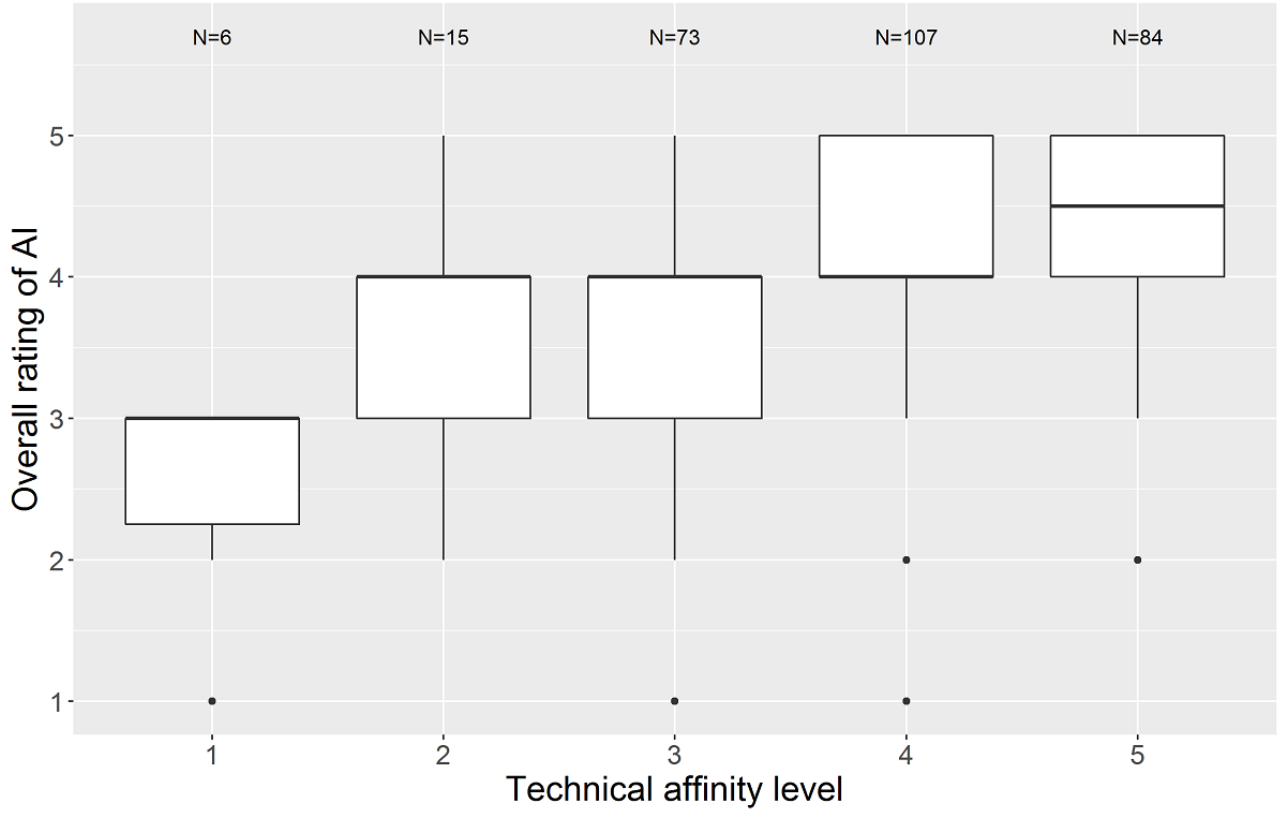
Figure 3. Effect of artificial intelligence (AI) on physicians' work.



The Kruskal-Wallis test revealed no significant difference in the personal rating of AI between the current occupations of assistant physician, medical specialist, senior physician, and clinic director (H_3=6.39, P=.09; see Multimedia Appendix 4). We could also not find a strong association between the AI score and the medical discipline groups described in the Methods section (H_5=5.92, P=.31; see Multimedia Appendix 5).

As expected, we found a significant association between the personal rating of AI in medicine and the self-reported technical affinity level (H_4=48.3, P<.001; Figure 4).

Figure 4. Personal rating of artificial intelligence (AI) stratified by technical affinity score.



AI in Medicine: Fields of Application With the Potential to Improve Clinical Practice

In the second section of the survey, we asked the physicians for their appraisal of the potential of AI in medicine to improve clinical care in various fields of application. In total, 25 AI applications were proposed. As described in the Methods section, we also stratified the applications for analysis into 6 categories (Textbox 1, Figure 5).

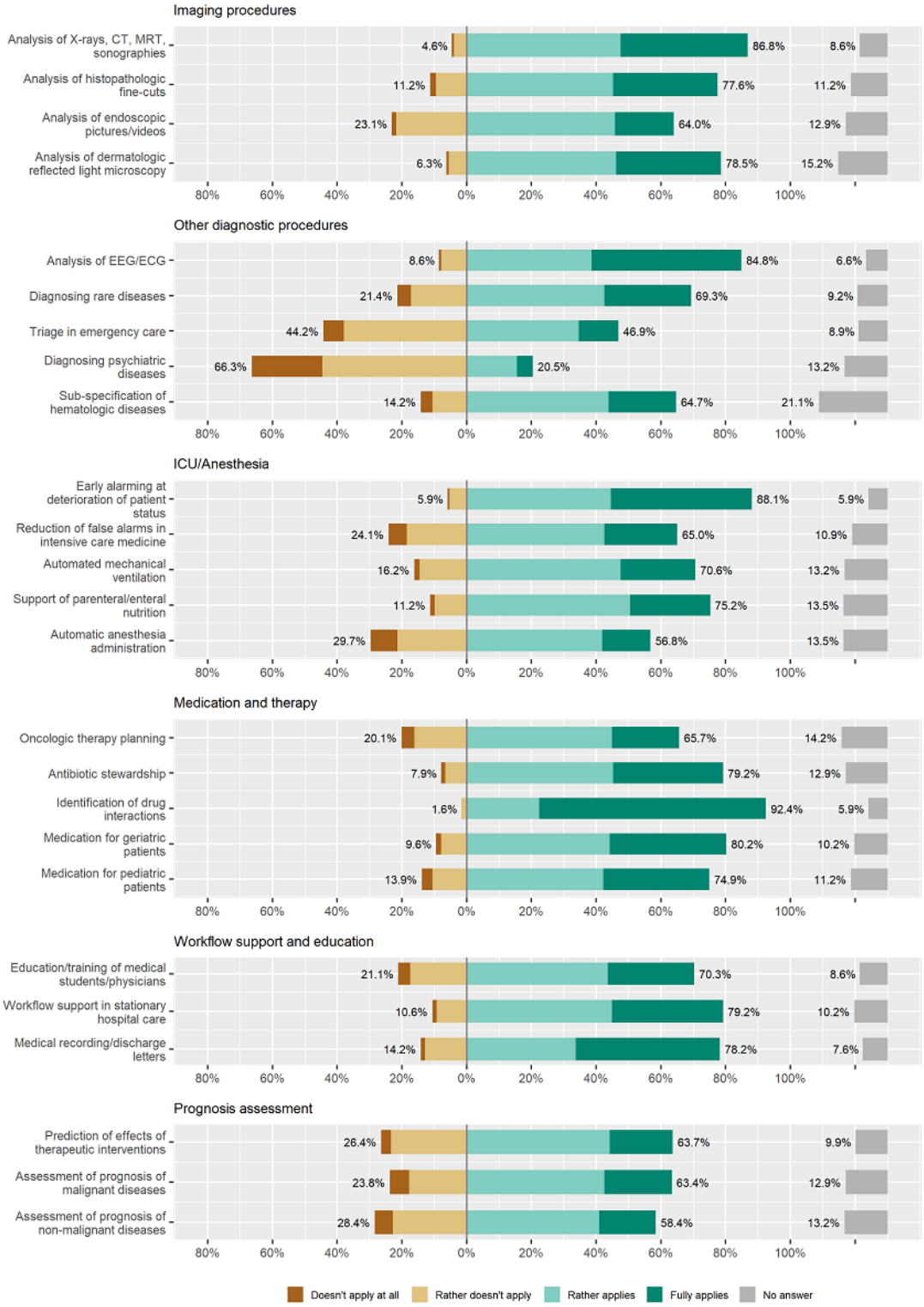
In the first category, “AI for imaging procedures,” a large majority of physicians agreed that all proposed applications had the potential to improve patient care substantially in the future. There was especially high agreement among respondents for the potential of AI to enhance the analysis of x-rays, computed tomography, magnetic resonance tomography, and sonographies (263/303, 86.8%). However, there was less agreement for the future potential of AI in the analysis of endoscopic images and videos (194/303, 64.0%) than for the other applications.

In the second category, “AI for other diagnostic procedures,” most physicians expected patient care to be improved significantly by using AI for the analysis of electroencephalograms and electrocardiograms and subspecification of hematologic diseases (257/303, 84.8%). Only a minority of respondents saw a role of AI in the diagnosis of psychiatric diseases (62/303, 20.5%) and in triage in emergency care (142/303, 46.9%).

The application of AI for ICU and anesthesia was assessed in the third category of AI applications for medicine. The majority of physicians agreed that all applications would improve patient care, even though the agreement for the potential of automatic anesthesia administration (172/303, 56.8%) was rated much lower than for the application of AI for an early alarm of the deterioration of patient status (267/303, 88.1%).

Furthermore, in the fourth and fifth categories “AI for medication and therapy” and “AI for workflow support and education,” respectively, a majority of respondents expected an improvement of daily practice through the listed AI applications. While AI's potential for the identification of drug interactions was outstandingly high (280/303, 92.4%), fewer physicians were convinced that AI for oncology therapy planning will advance patient care (199/303, 65.7%). Also, workflow support in stationary hospital care was expected and rated as beneficial for patient care in the future (240/303, 79.2%).

Figure 5. Artificial intelligence (AI) applications and potential for the future of medicine. CT: computed tomography; ECG: electrocardiogram; EEG: electroencephalogram; MRT: magnetic resonance tomography.



In the sixth category, regarding usage of AI for prognosis assessment, therapeutic interventions, and prognosis of malignant and nonmalignant diseases, we received lower agreement than in most other categories regarding an expected improvement in patient care.

In sum, in almost all categories of AI applications in medicine, physicians saw a high potential to improve patient care. A ranking of all proposed applications across all categories of applications with the highest number of answers for ”rather applies” and “fully applies” can be found in Table S2 in Multimedia Appendix 2. A short version, including the highest-rated applications and the lowest-rated potential to improve patient care in the future, is presented in Table 2. This table shows that the most frequently mentioned application to improve patient care was “Identification of drug interactions.” On the other side of the scale, the 2 least often mentioned AI

applications with potential for the future of health care were the usage of AI for "diagnosis of psychiatric diseases" and for "triage in emergency care." Beside those applications, there were also high numbers of "no responses" for specific applications like the application of AI for "Subspecification of hematologic diseases" (No response: 64/303, 21.1%).

Table 2. Applications with the highest-rated and lowest-rated potential to improve patient care in the future.

Rating	Artificial intelligence (AI) application	Field of application	Responses of "rather applies" or "fully applies", n (%)
1	Identification of drug interactions	Medication and therapy	280 (92.4)
2	Early alarming of deterioration of patient status	ICU[a]/anesthesia	267 (88.1)
3	Analysis of x-rays, CT[b], MRT[c], sonographies	Imaging procedures	263 (86.8)
4	Analysis of ECGs[d] and EEGs[e]	Other diagnostic procedures	257 (84.8)
[...]			
22	Assessment of prognosis of nonmalignant diseases	Prognosis assessment	177 (58.4)
23	Automatic anesthesia administration	ICU/anesthesia	172 (56.8)
24	Triage in emergency care	Other diagnostic procedures	142 (46.9)
25	Diagnosis of psychiatric diseases	Other diagnostic procedures	62 (20.5)

[a]ICU: intensive care unit.

[b]CT: computed tomography.

[c]MRT: magnetic resonance tomography.

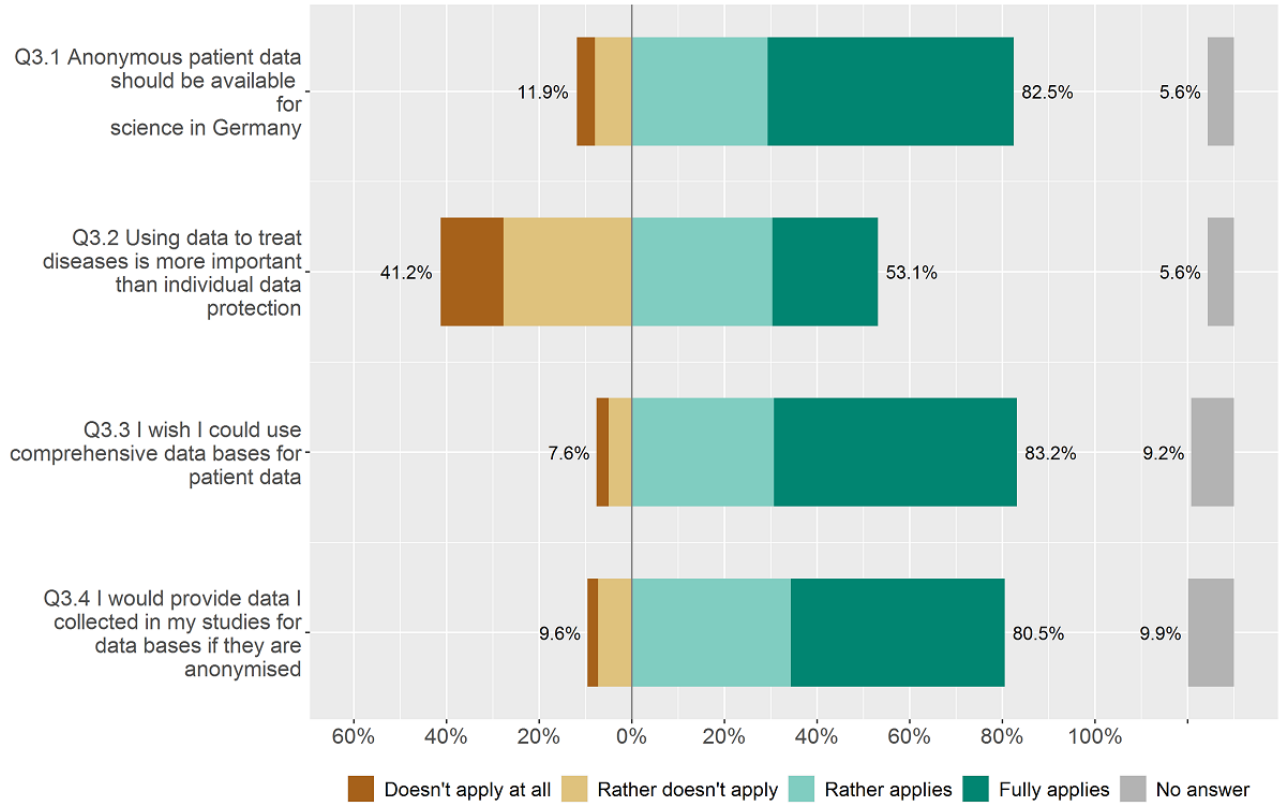
[d]ECG: electrocardiogram.

[e]EEG: electroencephalogram.

Personal Opinion of Physicians on the Secondary Usage of Anonymized Patient Data for Medical Research

In the third group of questions, we investigated physicians' attitudes on the secondary usage of anonymized patient data for medical data research, which in sum, was positive for the majority of respondents (see Figure 6). Solely for the question regarding balancing better treatment of diseases versus individual data protection, controversial answers were given. Most physicians thought that anonymized patient data should be freely available in Germany for research purposes (250/303, 82.5%), and 83.2% of respondents (252/303) agreed or fully agreed that they wished to be able to use anonymized patient data for their own research.

Figure 6. Attitude on the usage of (anonymized) patient data for medical research.



Discussion

Principal Findings

In our study, we could demonstrate that most physicians reported a positive or a very positive attitude towards AI in medicine. Physicians expect that AI will be used in clinical practice for various applications and will substantially improve patient care. There was agreement among the majority of physicians that AI will change their work as a physician. Participants also had a positive attitude towards using anonymized patient data for research purposes as a precondition for the development of algorithms for medical practice. Nevertheless, the usage of AI in medicine in today's clinical practice in hospitals and health care in Germany is rare.

Shaw et al [74] investigated the challenges of implementing AI in health care compared with other technologies, suggesting that AI implementation in medicine poses novel difficulties. Specific hurdles to be mastered are the absence of interoperabilitiy standards and missing regularities in cases of AI-driven, wrong decisions [23].

To overcome the problem of missing interoperability and improve access to health care data for clinical and biomedical research in Germany, the Ministry of Research and Education has initiated the German Medical Informatics Initiative to make clinical health data from patient care available for medical research [75]. Four consortia (Smart Medical Information Technology for Healthcare [SMITH] [76], Medical Informatics in Research and Care in University Medicine [MIRACUM] [77], HiGHmed [78], and Data Integration for Future Medicine [DIFUTURE] [79]) that include all German university hospitals are conceptualizing, developing, and operating so-called data integration centers in the university medical centers to make health care data from health information systems accessible for medical research. Beyond that, the initiative aims to create the regulatory framework and prerequisites for the secondary usage of routine health care data for (bio)medical data research [75].

Physicians' Attitudes Towards AI in Medicine

Physicians participating in our survey had a positive attitude towards the usage of AI in medicine. Nevertheless, they emphasized the need for scientific proof prior to broad implementation of AI-based systems. Besides the obligatory medical device instruction for AI in medicine, physicians want to have dedicated professional training to use AI in medicine. Another precondition for clinical usage of AI is clarity about the legal liability for its usage, especially when the basis of AI recommendations might not be easy to comprehend at once. Our survey's controversial answers to legal and regulatory questions show that rules and regulatory requirements are either not clear or nonexistent. Sullivan and Schweikart [29] described the complexity of legal responsibilities of health professionals and technology manufacturers, especially if the AI technology recommendations are not explainable. The authors highlighted one major problem: If the reasoning for recommendations is unknown, the AI is a black box, and there is a need for new legal solutions for AI usage in medicine.

While most physicians expect the future of medicine to be characterized by the combination of human and artificial intelligence, AI has already been proven to be able to outperform human physicians in specific tasks [80]. Nevertheless, human intelligence also learns from AI systems. As a restriction, the

authors think that an AI system making fully autonomous decisions would be neither desirable nor acceptable for the public.

Therefore, a hybrid solution of human and artificial intelligence can form a symbiotic relationship. Physicians expect that AI will significantly impact them and introduce changes for their daily work. This also includes the dependency of physicians on computer systems and new job requirements for physicians. Yu et al [81] described that AI can improve the quality of care by reducing human error and reducing human fatigue from a routine clinical task, but will probably not reduce physicians' workload, because medical guidelines might suggest higher frequencies of examinations for vulnerable patients. The authors Magrabi et al [28] described the challenges of evaluating AI-based decision support and AI's practical implications for medical practice. Due to the actual small number of AI applications, there is only little evidence to describe the concrete implications of AI for the clinical work of physicians. Nevertheless, it can be expected that analogously to the variety of AI applications in medical disciplines, physicians' work will change according to the task supported by the AI application. However, we see a general urgent need to integrate AI in medical education and professional training curricula [82,83].

Fields of AI Application With the Potential to Improve Clinical Practice

Currently, AI usage in medicine is one of the most promoted topics in medicine as a new technology that will fundamentally change physicians' clinical practice [32,84]. On the one hand, there are enormous expectations on AI-based decision support systems for better diagnosis, treatment, and clinical documentation facilitation. On the other hand, only a few AI applications have passed the regulatory requirements and have been implemented in clinical routine practice [85]. Participants in our study were optimistic that most proposed AI applications for medicine would improve patient care substantially in the future. The majority rated 23 of 25 applications positively, and 14 of these applications were evaluated by more than 70% of respondents to substantially enhance patient care. The highest potential was given for the AI application "identification of drug interactions," while the AI application for "diagnosis of psychiatric diseases" received the least positive evaluations. We assume that due to the increasing complexity of medication administration, physicians hope to be supported by AI-based decision support systems to avoid drug interactions especially when treating polymorbid, polypharmacy patients. The low rating of AI's potential for psychiatric diseases is remarkable as there are several AI applications for this medical field as described for ML in recent publications [52,53]. Respondents in the study population might not yet be completely informed about all potential AI applications, like in this context as well as the application of AI for speech or voice analysis using a recording.

Recently, Laguarta et al [86] from the Massachusetts Institute of Technology successfully developed and applied an AI model for diagnosing COVID-19 using only cough recordings, achieving a COVID-19 sensitivity of 98.5% with a specificity of 94.2% (area under the curve: 0.97). For asymptomatic subjects, the AI model achieved a sensitivity of 100% with a specificity of 83.2% [86]. The press release about the Massachusetts Institute of Technology Open Voice approach and the application of AI for voice analysis for the diagnosis of COVID-19 will contribute to informing physicians and the general population about less-known applications of AI for health care. To give physicians a basic understanding of AI for medical practice as well as its limitations and opportunities, Meskó and Görög [84] published a guide for medical professionals in the era of AI.

Table 2 shows that those applications using (sensor-based) continuously collected data in particular (eg, in ICUs or cardiology and neurology, imaging and video diagnostic procedures) and AI applications for workflow support were considered promising to substantially improve future patient care. In this context, Rush et al [87] argued that the data-rich ICU environment has massive AI usage potential. We can see that those fields were rated with high potential for future medicine improvement where information technology usage is high and structured data are documented. Less agreement among respondents was reached for fields with less structured data like therapeutic interviews in psychiatry.

Personal Opinion of Physicians on the Secondary Usage of Anonymized Patient Data for AI Development and Other Research Purposes

Access to clinical research databases like the MIMIC database is an elementary precondition for AI development for medicine. We found very positive attitudes towards the secondary usage of anonymized patient data for clinical and biomedical research. The fact that many researchers are positive about the anonymization and disclosure of their data after research can be a solution to the lack of publicly open medical data. Physicians' attitudes are in line with the general movement of science and engineering to make research data "findable, accessible, interoperable, and reusabile"—according to the FAIR data principles [88,89]. A European initiative applying the FAIR data principles is the European Open Science Cloud (EOSC), providing services to find and reuse each other's research objects under optimal and well-defined conditions [90]. The EOSC-Hub offers a wide selection of freeware services for researchers (eg, in the fields of data management, storage, data sharing, discovery, processing, and analysis) under one federated identity security system [90].

For the question addressing the trade-off between individual data protection and improvement of medical diagnosis or therapy, no unified opinion nor tendency was found. To avoid such issues, various privacy-preserving technologies have been developed for clinical and biomedical research, such as record linkage, synthetic data generation, and genomic data privacy [91]. Price and Cohen [92] described the legal and ethical challenges of big data for data privacy and how to handle patient data for the best conception of health privacy. The authors concluded that "Privacy underprotection and overprotection each create cognizable harms to patients both today and tomorrow," highlighting the enormous complexity of privacy in big data research.

Strengths and Limitations

We conducted a web-based survey among hospital physicians about their opinion of AI applications for different fields of applications and the attitudes of physicians towards the secondary usage of patient data for medical research. To our knowledge, this is the first survey to interrogate physicians' expectations and opinions of AI usage in medicine across German university hospitals. Yet, we did not investigate the usability of specific AI-based applications or decision support systems in our survey.

A limitation of our online survey is possible recruitment bias, as participating physicians may have had a particular interest or were involved in research on AI in medicine. The majority of respondents reported a positive attitude towards AI in health care. In addition, we found a positive association between self-reported technical affinity and attitude towards AI in medicine (Figure 4). Therefore, participants in our survey might have a more positive attitude towards technology and AI in health care than the whole population of physicians in German university hospitals. Even though physicians from almost every medical discipline participated in our survey, 42.2 % (128/303) were internal medicine and anesthesia department clinicians. This might have had an impact on the rating of the potential of AI application to improve health care in the future (Figure 5, Table 2). In consequence, AI applications performing tasks, which are common in these dominating disciplines, like identification of potential drug interactions and usage of continuous patient data monitoring could have received higher ratings than in a fully balanced population. Further studies on the usability and added value of AI applications in health care are needed prior to implemention in hospitals and medical practice in general. According to Magrabi et al [28], a rigorous initial and ongoing evaluation is essential for the safety and effectiveness of AI integration in sociotechnical settings like health care. Due to the huge variety and high complexity of AI applications, this paper may not take all regulatory issues into account.

Conclusions

Most physicians expect that medicine's future will be characterized by a combination of human and artificial intelligence. The participating physicians evaluated that most of the proposed AI applications will substantially improve patient care in the future. The highest potential is given to AI applications using sensor-based, continuously collected data like electrocardiogram and electroencephalogram or continuous patient monitoring in ICUs, imaging procedures in diagnostics, and workflow support. Physicians have the greatest expectation in the use of AI for the identification of drug interactions, reflecting the rising complexity of drug administration. Thus, future clinical AI users in hospitals seem to be ready for this new technology's clinical usage. We expect that AI applications will support imaging diagnostics and that AI applications for sensor-based, continuously collected data will be used in health care in the near future. In other medical disciplines with less standardization of data processing and collection, like in the German outpatient sector, AI applications will be developed and used in clinical practice later.

In general, the secondary usage of patient data and open access to databases for medical research were seen very positively by the physicians in our survey. Researchers in clinical and biomedical research would like to benefit from better access to research databases to generate new insights for improved patient care. Thus, initiatives like the German Medical Informatics Initiative, EOSC, and FAIR data principles will improve data usage from clinical care for clinical and (bio)medical research and facilitate researchers' access to clinical research data. In turn, that will fundamentally enhance the conditions of clinical data analysis and, as a consequence, enable better and personalized treatments for patients. Nevertheless, before new AI applications are implemented in clinical practice, the regulatory, legal, and ethical challenges must be mastered. Legislators and regulators must create the necessary framework for anonymous patient data exchange for clinical care and research, development of medical AI applications, and finally, its practical bedside use by physicians in health care.

Acknowledgments

This publication of the SMITH consortium was supported by the German Federal Ministry of Education and Research, grant numbers 01ZZ1803B, 01ZZ1803C, and 01ZZ1803M.

Authors' Contributions

OM and SF designed the survey and set up the web-based survey. OM, JP, and SF analyzed the data. OM wrote the manuscript. SF, JP, SD, JK, GM, MR, AS, and JB revised the article. All authors approved the final draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey on the usage of artificial intelligence in stationary hospital care.
[DOCX File , 36 KB - jmir_v23i3e26646_app1.docx]

Multimedia Appendix 2

Additional tables.
[DOCX File , 28 KB - jmir_v23i3e26646_app2.docx]

Multimedia Appendix 3
Personal rating of the usage of AI in medicine? (Scale: 1 – 5; 1 = very negative; 3 = neutral; 5 = very positive).
[PNG File , 26 KB - jmir_v23i3e26646_app3.png]

Multimedia Appendix 4
Overall rating of artificial intelligence (AI) stratified by current occupation.
[PNG File , 26 KB - jmir_v23i3e26646_app4.png]

Multimedia Appendix 5
Artificial intelligence (AI) affinity score stratified by medical discipline group.
[PNG File , 29 KB - jmir_v23i3e26646_app5.png]

References

1. Sanchez-Pinto LN, Luo Y, Churpek MM. Big Data and Data Science in Critical Care. Chest 2018 Nov;154(5):1239-1248 [FREE Full text] [doi: 10.1016/j.chest.2018.04.037] [Medline: 29752973]
2. Alexis Ruiz A, Wyszyńska PK, Laudanski K. Narrative Review of Decision-Making Processes in Critical Care. Anesthesia & Analgesia 2019 May;128(5):962-970. [doi: 10.1213/ANE.0000000000003683] [Medline: 30096084]
3. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial Intelligence in Health Care: Bibliometric Analysis. J Med Internet Res 2020 Jul 29;22(7):e18228 [FREE Full text] [doi: 10.2196/18228] [Medline: 32723713]
4. Rong G, Mendez A, Bou Assi E, Zhao B, Sawan M. Artificial Intelligence in Healthcare: Review and Prediction Case Studies. Engineering 2020 Mar;6(3):291-301 [FREE Full text] [doi: 10.1016/j.eng.2019.08.015]
5. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. JMLR Workshop Conf Proc 2016 Aug;56:301-318 [FREE Full text] [Medline: 28286600]
6. Fagherazzi G, Ravaud P. Digital diabetes: Perspectives for diabetes prevention, management and research. Diabetes Metab 2019 Sep;45(4):322-329. [doi: 10.1016/j.diabet.2018.08.012] [Medline: 30243616]
7. Wang L, Lin ZQ, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Sci Rep 2020 Nov 11;10(1):19549 [FREE Full text] [doi: 10.1038/s41598-020-76550-z] [Medline: 33177550]
8. Fröhlich H, Balling R, Beerenwinkel N, Kohlbacher O, Kumar S, Lengauer T, et al. From hype to reality: data science enabling personalized medicine. BMC Med 2018 Aug 27;16(1):150 [FREE Full text] [doi: 10.1186/s12916-018-1122-7] [Medline: 30145981]
9. Kumar M, Vimala R, Britto K. A cognitive technology based healthcare monitoring system and medical data transmission. Measurement 2019 Nov;146:322-332 [FREE Full text] [doi: 10.1016/j.measurement.2019.03.017]
10. Organizing Committee of the Madrid 2017 Critical Care Datathon, Núñez Reiz A, Martínez Sagasti F, Álvarez González M, Blesa Malpica A, Martín Benítez JC, et al. Big data and machine learning in critical care: Opportunities for collaborative research. Med Intensiva 2019;43(1):52-57 [FREE Full text] [doi: 10.1016/j.medin.2018.06.002] [Medline: 30077427]
11. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
12. Chen D, Liu S, Kingsbury P, Sohn S, Storlie CB, Habermann EB, et al. Deep learning and alternative learning strategies for retrospective real-world clinical data. NPJ Digit Med 2019;2:43 [FREE Full text] [doi: 10.1038/s41746-019-0122-0] [Medline: 31304389]
13. Montazeri M, Montazeri M, Montazeri M, Beigzadeh A. Machine learning models in breast cancer survival prediction. Technol Health Care 2016;24(1):31-42. [doi: 10.3233/THC-151071] [Medline: 26409558]
14. Hosny A, Parmar C, Coroller TP, Grossmann P, Zeleznik R, Kumar A, et al. Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. PLoS Med 2018 Nov;15(11):e1002711 [FREE Full text] [doi: 10.1371/journal.pmed.1002711] [Medline: 30500819]
15. Kather JN, Krisam J, Charoentong P, Luedde T, Herpel E, Weis C, et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. PLoS Med 2019 Jan;16(1):e1002730 [FREE Full text] [doi: 10.1371/journal.pmed.1002730] [Medline: 30677016]
16. Parreco J, Hidalgo A, Kozol R, Namias N, Rattan R. Predicting Mortality in the Surgical Intensive Care Unit Using Artificial Intelligence and Natural Language Processing of Physician Documentation. Am Surg 2018 Jul 01;84(7):1190-1194. [Medline: 30064586]
17. Meiring C, Dixit A, Harris S, MacCallum NS, Brealey DA, Watkinson PJ, et al. Optimal intensive care outcome prediction over time using machine learning. PLoS One 2018;13(11):e0206862 [FREE Full text] [doi: 10.1371/journal.pone.0206862] [Medline: 30427913]

18. Xu F, Sepúlveda MJ, Jiang Z, Wang H, Li J, Liu Z, et al. Effect of an Artificial Intelligence Clinical Decision Support System on Treatment Decisions for Complex Breast Cancer. JCO Clin Cancer Inform 2020 Sep;4:824-838 [FREE Full text] [doi: 10.1200/CCI.20.00018] [Medline: 32970484]
19. Baltruschat I, Steinmeister L, Nickisch H, Saalbach A, Grass M, Adam G, et al. Smart chest X-ray worklist prioritization using artificial intelligence: a clinical workflow simulation. Eur Radiol 2020 Nov 21:1. [doi: 10.1007/s00330-020-07480-7] [Medline: 33219850]
20. Rush B, Celi LA, Stone DJ. Applying machine learning to continuously monitored physiological data. J Clin Monit Comput 2019 Oct;33(5):887-893 [FREE Full text] [doi: 10.1007/s10877-018-0219-z] [Medline: 30417258]
21. Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
22. Article metrics: MIMIC-III, a freely accessible critical care database. Scientific data. 2021. URL: https://www.nature.com/articles/sdata201635/metrics [accessed 2021-02-21]
23. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
24. Willemink MJ, Noël PB. The evolution of image reconstruction for CT-from filtered back projection to artificial intelligence. Eur Radiol 2019 May;29(5):2185-2195 [FREE Full text] [doi: 10.1007/s00330-018-5810-7] [Medline: 30377791]
25. Manz CR, Chen J, Liu M, Chivers C, Regli SH, Braun J, et al. Validation of a Machine Learning Algorithm to Predict 180-Day Mortality for Outpatients With Cancer. JAMA Oncol 2020 Sep 24:1. [doi: 10.1001/jamaoncol.2020.4331] [Medline: 32970131]
26. Brajer N, Cozzi B, Gao M, Nichols M, Revoir M, Balu S, et al. Prospective and External Evaluation of a Machine Learning Model to Predict In-Hospital Mortality of Adults at Time of Admission. JAMA Netw Open 2020 Feb 05;3(2):e1920733 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.20733] [Medline: 32031645]
27. Cabitza F, Campagner A, Balsano C. Bridging the "last mile" gap between AI implementation and operation: "data awareness" that matters. Ann Transl Med 2020 Apr;8(7):501 [FREE Full text] [doi: 10.21037/atm.2020.03.63] [Medline: 32395545]
28. Magrabi F, Ammenwerth E, McNair JB, De Keizer NF, Hyppönen H, Nykänen P, et al. Artificial Intelligence in Clinical Decision Support: Challenges for Evaluating AI and Practical Implications. Yearb Med Inform 2019 Aug;28(1):128-134 [FREE Full text] [doi: 10.1055/s-0039-1677903] [Medline: 31022752]
29. Sullivan HR, Schweikart SJ. Are Current Tort Liability Doctrines Adequate for Addressing Injury Caused by AI? AMA J Ethics 2019 Feb 01;21(2):E160-E166 [FREE Full text] [doi: 10.1001/amajethics.2019.160] [Medline: 30794126]
30. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digit Med 2020;3:119 [FREE Full text] [doi: 10.1038/s41746-020-00323-1] [Medline: 33015372]
31. Raghu M, Zhang C, Kleinberg J, Bengio S. Transfusion: Understanding transfer learning for medical imaging. Advances in Neural Information Processing Systems 32 (NeurIPS 2019). 2019. URL: https://papers.nips.cc/paper/2019/hash/eb1e78328c46506b46a4ac4a1e378b91-Abstract.html [accessed 2021-02-21]
32. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
33. Lüdecke D. sjPlot: Data Visualization for Statistics in Social Science. R package version 2. URL: https://CRAN.R-project.org/package=sjPlot [accessed 2021-02-21]
34. Wickham H, Navarro D, Pedersen TL. ggplot2: elegant graphics for data analysis. New York, NY: Springer-Verlag; 2016.
35. R-Core-Team. R: A language and environment for statistical computing (Version 4.0.3). The R Foundation. URL: https://www.R-project.org [accessed 2021-02-21]
36. Wan Y, Wu PW, Huang P, Tsay P, Pan K, Trang NN, et al. The Use of Artificial Intelligence in the Differentiation of Malignant and Benign Lung Nodules on Computed Tomograms Proven by Surgical Pathology. Cancers (Basel) 2020 Aug 07;12(8):1 [FREE Full text] [doi: 10.3390/cancers12082211] [Medline: 32784681]
37. Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S, et al. Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci U S A 2018 Nov 06;115(45):11591-11596 [FREE Full text] [doi: 10.1073/pnas.1806905115] [Medline: 30348771]
38. Titano JJ, Badgeley M, Schefflein J, Pain M, Su A, Cai M, et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. Nat Med 2018 Sep;24(9):1337-1341. [doi: 10.1038/s41591-018-0147-y] [Medline: 30104767]
39. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, the CAMELYON16 Consortium, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA 2017 Dec 12;318(22):2199-2210 [FREE Full text] [doi: 10.1001/jama.2017.14585] [Medline: 29234806]
40. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med 2018 Oct;24(10):1559-1567. [doi: 10.1038/s41591-018-0177-5] [Medline: 30224757]

41. Misawa M, Kudo S, Mori Y, Cho T, Kataoka S, Yamauchi A, et al. Artificial Intelligence-Assisted Polyp Detection for Colonoscopy: Initial Experience. Gastroenterology 2018 Jun;154(8):2027-2029.e3 [FREE Full text] [doi: 10.1053/j.gastro.2018.04.003] [Medline: 29653147]
42. Urban G, Tripathi P, Alkayali T, Mittal M, Jalali F, Karnes W, et al. Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. Gastroenterology 2018 Oct;155(4):1069-1078.e8 [FREE Full text] [doi: 10.1053/j.gastro.2018.06.037] [Medline: 29928897]
43. Alagappan M, Brown JRG, Mori Y, Berzin TM. Artificial intelligence in gastrointestinal endoscopy: The future is almost here. World J Gastrointest Endosc 2018 Oct 16;10(10):239-249 [FREE Full text] [doi: 10.4253/wjge.v10.i10.239] [Medline: 30364792]
44. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Jan 25;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
45. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Reader study level-Ilevel-II Groups, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol 2018 Aug 01;29(8):1836-1842 [FREE Full text] [doi: 10.1093/annonc/mdy166] [Medline: 29846502]
46. Rajagopalan SS, Bhardwaj S, Panda R, Reddam VR, Ganne C, Kenchaiah R, et al. Machine learning detects EEG microstate alterations in patients living with temporal lobe epilepsy. Seizure 2018 Oct;61:8-13. [doi: 10.1016/j.seizure.2018.07.007] [Medline: 30044996]
47. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019 Jan;25(1):65-69 [FREE Full text] [doi: 10.1038/s41591-018-0268-3] [Medline: 30617320]
48. Kiely DG, Doyle O, Drage E, Jenner H, Salvatelli V, Daniels FA, et al. Utilising artificial intelligence to determine patients at risk of a rare disease: idiopathic pulmonary arterial hypertension. Pulm Circ 2019;9(4):2045894019890549 [FREE Full text] [doi: 10.1177/2045894019890549] [Medline: 31798836]
49. Grigull L, Mehmecke S, Rother A, Blöß S, Klemann C, Schumacher U, et al. Common pre-diagnostic features in individuals with different rare diseases represent a key for diagnostic support with computerized pattern recognition? PLoS One 2019;14(10):e0222637 [FREE Full text] [doi: 10.1371/journal.pone.0222637] [Medline: 31600214]
50. Kang D, Cho K, Kwon O, Kwon J, Jeon K, Park H, et al. Artificial intelligence algorithm to predict the need for critical care in prehospital emergency medical services. Scand J Trauma Resusc Emerg Med 2020 Mar 04;28(1):17 [FREE Full text] [doi: 10.1186/s13049-020-0713-4] [Medline: 32131867]
51. Tahayori B, Chini-Foroush N, Akhlaghi H. Advanced natural language processing technique to predict patient disposition based on emergency triage notes. Emerg Med Australas 2020 Oct 11:1. [doi: 10.1111/1742-6723.13656] [Medline: 33043570]
52. Aafjes-van Doorn K, Kamsteeg C, Bate J, Aafjes M. A scoping review of machine learning in psychotherapy research. Psychother Res 2021 Jan;31(1):92-116. [doi: 10.1080/10503307.2020.1808729] [Medline: 32862761]
53. Fiske A, Henningsen P, Buyx A. Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. J Med Internet Res 2019 May 09;21(5):e13216 [FREE Full text] [doi: 10.2196/13216] [Medline: 31094356]
54. Radakovich N, Cortese M, Nazha A. Acute myeloid leukemia and artificial intelligence, algorithms and new scores. Best Pract Res Clin Haematol 2020 Sep;33(3):101192 [FREE Full text] [doi: 10.1016/j.beha.2020.101192] [Medline: 33038981]
55. Warnat-Herresthal S, Perrakis K, Taschler B, Becker M, Baßler K, Beyer M, et al. Scalable Prediction of Acute Myeloid Leukemia Using High-Dimensional Machine Learning and Blood Transcriptomics. iScience 2020 Jan 24;23(1):100780 [FREE Full text] [doi: 10.1016/j.isci.2019.100780] [Medline: 31918046]
56. Kratz A, Lee S, Zini G, Riedl JA, Hur M, Machin S, International Council for Standardization in Haematology. Digital morphology analyzers in hematology: ICSH review and recommendations. Int J Lab Hematol 2019 Aug;41(4):437-447. [doi: 10.1111/ijlh.13042] [Medline: 31046197]
57. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. Crit Care Med 2016 Feb;44(2):368-374 [FREE Full text] [doi: 10.1097/CCM.0000000000001571] [Medline: 26771782]
58. Chen L, Dubrawski A, Wang D, Fiterau M, Guillame-Bert M, Bose E, et al. Using Supervised Machine Learning to Classify Real Alerts and Artifact in Online Multisignal Vital Sign Monitoring Data. Crit Care Med 2016 Jul;44(7):e456-e463 [FREE Full text] [doi: 10.1097/CCM.0000000000001660] [Medline: 26992068]
59. Sottile PD, Albers D, Higgins C, Mckeehan J, Moss MM. The Association Between Ventilator Dyssynchrony, Delivered Tidal Volume, and Sedation Using a Novel Automated Ventilator Dyssynchrony Detection Algorithm. Crit Care Med 2018 Feb;46(2):e151-e157 [FREE Full text] [doi: 10.1097/CCM.0000000000002849] [Medline: 29337804]
60. Schäublin J, Derighetti M, Feigenwinter P, Petersen-Felix S, Zbinden AM. Fuzzy logic control of mechanical ventilation during anaesthesia. Br J Anaesth 1996 Nov;77(5):636-641 [FREE Full text] [doi: 10.1093/bja/77.5.636] [Medline: 8957981]
61. Kreif N, Tran L, Grieve R, De Stavola B, Tasker RC, Petersen M. Estimating the Comparative Effectiveness of Feeding Interventions in the Pediatric Intensive Care Unit: A Demonstration of Longitudinal Targeted Maximum Likelihood

Estimation. Am J Epidemiol 2017 Dec 15;186(12):1370-1379 [FREE Full text] [doi: 10.1093/aje/kwx213] [Medline: 28992064]
62. Lee H, Ryu H, Chung E, Jung C. Prediction of Bispectral Index during Target-controlled Infusion of Propofol and Remifentanil: A Deep Learning Approach. Anesthesiology 2018 Mar;128(3):492-501 [FREE Full text] [doi: 10.1097/ALN.0000000000001892] [Medline: 28953500]
63. Zhao X, Zhang Y, Ma X, Chen Y, Xi J, Yin X, et al. Concordance between treatment recommendations provided by IBM Watson for Oncology and a multidisciplinary tumor board for breast cancer in China. Jpn J Clin Oncol 2020 Aug 04;50(8):852-858. [doi: 10.1093/jjco/hyaa051] [Medline: 32419014]
64. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. Crit Care Med 2018 Apr;46(4):547-553 [FREE Full text] [doi: 10.1097/CCM.0000000000002936] [Medline: 29286945]
65. Deng Y, Xu X, Qiu Y, Xia J, Zhang W, Liu S. A multimodal deep learning framework for predicting drug-drug interaction events. Bioinformatics 2020 Aug 01;36(15):4316-4322. [doi: 10.1093/bioinformatics/btaa501] [Medline: 32407508]
66. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics 2018 Jul 01;34(13):i457-i466 [FREE Full text] [doi: 10.1093/bioinformatics/bty294] [Medline: 29949996]
67. Liu N, Chen C, Kumara S. Semi-Supervised Learning Algorithm for Identifying High-Priority Drug-Drug Interactions Through Adverse Event Reports. IEEE J Biomed Health Inform 2020 Jan;24(1):57-68. [doi: 10.1109/JBHI.2019.2932740] [Medline: 31395567]
68. Sessa M, Rasmussen DB, Jensen MT, Kragholm K, Torp-Pedersen C, Andersen M. Metoprolol Versus Carvedilol in Patients With Heart Failure, Chronic Obstructive Pulmonary Disease, Diabetes Mellitus, and Renal Failure. Am J Cardiol 2020 Apr 01;125(7):1069-1076. [doi: 10.1016/j.amjcard.2019.12.048] [Medline: 32000982]
69. Fanelli U, Pappalardo M, Chinè V, Gismondi P, Neglia C, Argentiero A, et al. Role of Artificial Intelligence in Fighting Antimicrobial Resistance in Pediatrics. Antibiotics (Basel) 2020 Nov 01;9(11):1 [FREE Full text] [doi: 10.3390/antibiotics9110767] [Medline: 33139605]
70. Mirchi N, Bissonnette V, Yilmaz R, Ledwos N, Winkler-Schwartz A, Del Maestro RF. The Virtual Operative Assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine. PLoS One 2020;15(2):e0229596 [FREE Full text] [doi: 10.1371/journal.pone.0229596] [Medline: 32106247]
71. Lin SY, Shanafelt TD, Asch SM. Reimagining Clinical Documentation With Artificial Intelligence. Mayo Clin Proc 2018 May;93(5):563-565. [doi: 10.1016/j.mayocp.2018.02.016] [Medline: 29631808]
72. Coiera E, Kocaballi B, Halamka J, Laranjo L. The digital scribe. npj Digital Med 2018 Oct 16;1(1):58 [FREE Full text] [doi: 10.1038/s41746-018-0066-9] [Medline: 31304337]
73. Menden MP, Iorio F, Garnett M, McDermott U, Benes CH, Ballester PJ, et al. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. PLoS One 2013;8(4):e61318 [FREE Full text] [doi: 10.1371/journal.pone.0061318] [Medline: 23646105]
74. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial Intelligence and the Implementation Challenge. J Med Internet Res 2019 Jul 10;21(7):e13659 [FREE Full text] [doi: 10.2196/13659] [Medline: 31293245]
75. Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. Methods Inf Med 2018 Jul;57(S 01):e50-e56 [FREE Full text] [doi: 10.3414/ME18-03-0003] [Medline: 30016818]
76. Winter A, Stäubert S, Ammon D, Aiche S, Beyan O, Bischoff V, et al. Smart Medical Information Technology for Healthcare (SMITH). Methods Inf Med 2018 Jul;57(S 01):e92-e105 [FREE Full text] [doi: 10.3414/ME18-02-0004] [Medline: 30016815]
77. Prokosch H, Acker T, Bernarding J, Binder H, Boeker M, Boerries M, et al. MIRACUM: Medical Informatics in Research and Care in University Medicine. Methods Inf Med 2018 Jul;57(S 01):e82-e91 [FREE Full text] [doi: 10.3414/ME17-02-0025] [Medline: 30016814]
78. Haarbrandt B, Schreiweis B, Rey S, Sax U, Scheithauer S, Rienhoff O, et al. HiGHmed - An Open Platform Approach to Enhance Care and Research across Institutional Boundaries. Methods Inf Med 2018 Jul;57(S 01):e66-e81 [FREE Full text] [doi: 10.3414/ME18-02-0002] [Medline: 30016813]
79. Prasser F, Kohlbacher O, Mansmann U, Bauer B, Kuhn KA. Data Integration for Future Medicine (DIFUTURE). Methods Inf Med 2018 Jul;57(S 01):e57-e65 [FREE Full text] [doi: 10.3414/ME17-02-0022] [Medline: 30016812]
80. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 2018 Nov;24(11):1716-1720. [doi: 10.1038/s41591-018-0213-5] [Medline: 30349085]
81. Yu K, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018 Oct;2(10):719-731 [FREE Full text] [doi: 10.1038/s41551-018-0305-z] [Medline: 31015651]
82. Lindqwister AL, Hassanpour S, Lewis PJ, Sin JM. AI-RADS: An Artificial Intelligence Curriculum for Residents. Acad Radiol 2020 Oct 15:1 [FREE Full text] [doi: 10.1016/j.acra.2020.09.017] [Medline: 33071185]
83. Sapci AH, Sapci HA. Artificial Intelligence Education and Tools for Medical and Health Informatics Students: Systematic Review. JMIR Med Educ 2020 Jun 30;6(1):e19285 [FREE Full text] [doi: 10.2196/19285] [Medline: 32602844]

84. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. NPJ Digit Med 2020;3:126 [FREE Full text] [doi: 10.1038/s41746-020-00333-z] [Medline: 33043150]
85. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med 2020;3:118 [FREE Full text] [doi: 10.1038/s41746-020-00324-0] [Medline: 32984550]
86. Laguarta J, Hueto F, Subirana B. COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings. IEEE Open J. Eng. Med. Biol 2020;1:275-281. [doi: 10.1109/ojemb.2020.3026928]
87. Rush B, Stone DJ, Celi LA. From Big Data to Artificial Intelligence: Harnessing Data Routinely Collected in the Process of Care. Crit Care Med 2018 Feb;46(2):345-346 [FREE Full text] [doi: 10.1097/CCM.0000000000002892] [Medline: 29337803]
88. Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data. European Commission. 2018. URL: https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf [accessed 2021-02-21]
89. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3:160018 [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
90. Directorate-General for Research and Innovation (European Commission). Realising the european open science cloud. Publications Office of the European Union. 2016. URL: https://op.europa.eu/en/publication-detail/-/publication/2ec2eced-9ac5-11e6-868c-01aa75ed71a1/language-en [accessed 2021-02-21]
91. Wang S, Bonomi L, Dai W, Chen F, Cheung C, Bloss CS, et al. Big Data Privacy in Biomedical Research. IEEE Trans Big Data 2020 Jun;6(2):296-308. [doi: 10.1109/TBDATA.2016.2608848] [Medline: 32478127]
92. Price WN, Cohen IG. Privacy in the age of medical big data. Nat Med 2019 Jan;25(1):37-43 [FREE Full text] [doi: 10.1038/s41591-018-0272-7] [Medline: 30617331]

Abbreviations

AI: artificial intelligence
DIFUTURE: Data Integration for Future Medicine
EOSC: European Open Science Cloud
ICU: intensive care unit
MIMIC: Medical Information Mart for Intensive Care
MIRACUM: Medical Informatics in Research and Care in University Medicine
ML: machine learning
SMITH: Smart Medical Information Technology for Healthcare

Edited by G Eysenbach; submitted 19.12.20; peer-reviewed by HC Lee; comments to author 13.01.21; revised version received 29.01.21; accepted 15.02.21; published 05.03.21.

Please cite as:
Maassen O, Fritsch S, Palm J, Deffge S, Kunze J, Marx G, Riedel M, Schuppert A, Bickenbach J
Future Medical Artificial Intelligence Application Requirements and Expectations of Physicians in German University Hospitals: Web-Based Survey
J Med Internet Res 2021;23(3):e26646
URL: https://www.jmir.org/2021/3/e26646
doi:10.2196/26646
PMID:33666563

Original Paper

Natural Language Processing and Machine Learning for Identifying Incident Stroke From Electronic Health Records: Algorithm Development and Validation

Yiqing Zhao[1], PhD; Sunyang Fu[1], MHI; Suzette J Bielinski[1], PhD; Paul A Decker[1], MSc; Alanna M Chamberlain[1], PhD; Veronique L Roger[1], MD; Hongfang Liu[1], PhD; Nicholas B Larson[1], PhD

Department of Health Sciences Research, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Nicholas B Larson, PhD
Department of Health Sciences Research
Mayo Clinic
205 3rd Ave SW
Rochester, MN, 55905
United States
Phone: 1 507 293 1700
Email: Larson.Nicholas@mayo.edu

Abstract

Background: Stroke is an important clinical outcome in cardiovascular research. However, the ascertainment of incident stroke is typically accomplished via time-consuming manual chart abstraction. Current phenotyping efforts using electronic health records for stroke focus on case ascertainment rather than incident disease, which requires knowledge of the temporal sequence of events.

Objective: The aim of this study was to develop a machine learning–based phenotyping algorithm for incident stroke ascertainment based on diagnosis codes, procedure codes, and clinical concepts extracted from clinical notes using natural language processing.

Methods: The algorithm was trained and validated using an existing epidemiology cohort consisting of 4914 patients with atrial fibrillation (AF) with manually curated incident stroke events. Various combinations of feature sets and machine learning classifiers were compared. Using a heuristic rule based on the composition of concepts and codes, we further detected the stroke subtype (ischemic stroke/transient ischemic attack or hemorrhagic stroke) of each identified stroke. The algorithm was further validated using a cohort (n=150) stratified sampled from a population in Olmsted County, Minnesota (N=74,314).

Results: Among the 4914 patients with AF, 740 had validated incident stroke events. The best-performing stroke phenotyping algorithm used clinical concepts, diagnosis codes, and procedure codes as features in a random forest classifier. Among patients with stroke codes in the general population sample, the best-performing model achieved a positive predictive value of 86% (43/50; 95% CI 0.74-0.93) and a negative predictive value of 96% (96/100). For subtype identification, we achieved an accuracy of 83% in the AF cohort and 80% in the general population sample.

Conclusions: We developed and validated a machine learning–based algorithm that performed well for identifying incident stroke and for determining type of stroke. The algorithm also performed well on a sample from a general population, further demonstrating its generalizability and potential for adoption by other institutions.

(J Med Internet Res 2021;23(3):e22951) doi:10.2196/22951

KEYWORDS

stroke; natural language processing; electronic health records; machine learning

Introduction

Stroke is a syndrome involving a rapid loss of cerebral function with vascular origin [1]. The loss of function can result in deep coma or subarachnoid hemorrhage. There are two broad categories of stroke: hemorrhagic and ischemic stroke [2]. Hemorrhage is caused by bleeding within the skull cavity, while ischemia is characterized by inadequate blood to supply a part of the brain. Stroke identification is an important outcome for various cardiovascular studies [3-5]. However, a challenge with stroke ascertainment is the inconsistent use of International

Classification of Diseases (ICD) codes [6], which may result in inaccurate code-based ascertainment of cases [7]. Therefore, the time-consuming process of electronic health record (EHR) abstraction remains the gold standard of stroke ascertainment [8,9].

Machine learning has recently gained popularity for its ability to classify patients or make predictions on various aspects of diseases. In contrast to manually curated algorithms based on domain expertise, machine learning is a data-driven approach that can be trained on large data sets to identify and leverage complex feature relationships and improve classification and prediction tasks thereby. In terms of stroke, machine learning algorithms have been applied to predict future stroke cases [10], mortality and recurrent strokes [11,12], and treatment outcomes [13,14]. Most existing phenotyping algorithms have been developed to only differentiate between cases and noncases of diseases [15-18]; however, ascertaining incident disease (ie, first occurrence of disease) in a population is a more difficult task [8,19,20]. A recent study by Ni et al [21] examined potential predictive features of stroke occurrence including demographic, clinical, and diagnostic characteristics of patients. The authors found that diagnostic tests for stroke, such as computed tomography (CT) and magnetic resonance imaging (MRI), contributed to most of the model performance, and that the optimal feature set included imaging findings, signs and symptoms, interventions, emergency department assessments, findings from angiography and carotid ultrasound tests, ICD codes, substance use (smoking, alcohol, and street drugs) characteristics, and demographics. However, features such as signs and symptoms, substance use characteristics, and demographics may not be specific enough for disease ascertainment, as there is a high prevalence of strokelike symptoms among people without a diagnosis of stroke [22]. In addition, incorporating too many features in the model may result in overfitting without appropriate regularization. Another study [7] also used ICD and Current Procedural Terminology (CPT) [23] codes as features to classify positive, possible, and negative stroke cases. However, stroke-related clinical concepts (including both disease name concepts and symptom concepts) in unstructured clinical notes were not included in this model.

Rapid adoption of EHRs has enabled secondary use of the EHR data in epidemiological research [24-26]. Previous studies noted the existence of bias using a single type of EHR data (ie, diagnosis codes) [27-29]. To avoid this bias, the Electronic Medical Records and Genomics (eMERGE) consortium [30,31] has piloted the development of EHR-based phenotyping algorithms using multiple types of EHR data [32-34]. This has given rise to a number of phenotyping algorithms that use both structured EHR data (eg, demographics, diagnosis and procedure codes, laboratory test results, and medications) and unstructured EHR data (eg, clinical notes, imaging reports, and discharge summaries) [35-38]. However, the eMERGE consortium algorithms are typically focused on identifying cases and noncases rather than characterizing a new-onset (ie, incident) disease in a population. Moreover, extracting information from unstructured clinical text is a nontrivial task that involves natural language processing techniques [39-41].

In our paper, we address existing challenges for stroke ascertainment, specifically for incident stroke. Our research objective is to develop and validate a machine learning–based phenotyping algorithm to identify incident stroke and detailed stroke subtypes based on three major EHR-derived data elements: clinical concepts extracted from clinical notes; ICD, Ninth Revision (ICD-9) diagnosis codes; and CPT procedure codes.

Methods

This study was approved by the Mayo Clinic Institutional Review Board (no. 17-008818) and is in accordance with the ethical standards mandated by the committee on responsible human experimentation. The data that support the findings of this study are available from the corresponding author upon reasonable request.

Study Design

This was a predictive modeling study that used observational cohort data for training and validation. We employed an atrial fibrillation (AF) cohort, in which all incidences of stroke were manually ascertained in a previous study [4], to train and test our phenotyping algorithm for the date of incident stroke events. We then evaluated the generalizability of our algorithm in a general population cohort.

The AF Cohort

The AF cohort comprised a patient population from Olmsted County, Minnesota, USA [4,42]. Olmsted County is an area relatively isolated from other urban centers with only a few providers delivering most care to residents, primarily Mayo Clinic and Olmsted Medical Center [43-45]. Extracting all health care–related events was completed through the Rochester Epidemiology Project (REP), a records linkage system [43,44]. The REP is a records linkage system that allows retrieval of nearly all health care utilization and outcomes of residents living in Olmsted County. The electronic indexes of the REP include demographic information, diagnostic and procedure codes, health care utilization data, outpatient drug prescriptions, results of laboratory tests, and information about smoking, height, weight, and body mass index. ICD-9 codes and the Mayo Clinic electrocardiograms were obtained among adults aged ≥18 years from 2000 to 2014 to ascertain AF. Patients were identified by the presence of an ICD-9 code for stroke through March 31, 2015, and then validated by manual review of the EHR. Strokes were classified as ischemic strokes/transient ischemic attack or hemorrhagic strokes [4,46]. The first (incident) event of each type of stroke after the incident AF date was ascertained, regardless of whether a patient had a prior stroke. The AF cohort included 4914 validated patients with AF, 1773 of whom were screened for a possible stroke. Table 1 shows the cohort characteristics. Manual abstraction of the EHR validated the stroke code in 740 patients. Manual ascertainment of stroke and the dates of the events were used as a gold standard to train and test the stroke algorithm.

Table 1. Atrial fibrillation cohort characteristics.

Measure	Cohort (n=4914)	Screened (n=1773)
Gender, n (%)		
Female	2309 (46.99)	869 (49.01)
Male	2605 (53.01)	904 (50.99)
Age at diagnosis of AF[a] (years), mean		
Female	76	80
Male	70	74
ICD-9[b] diagnosis codes[c], n	27,243	27,243

[a]AF: atrial fibrillation.

[b]ICD-9: International Classification of Diseases, Ninth Revision.

[c]ICD retrieval was from AF incidence date to March 31, 2015. AF validations were from 2000 to 2014.

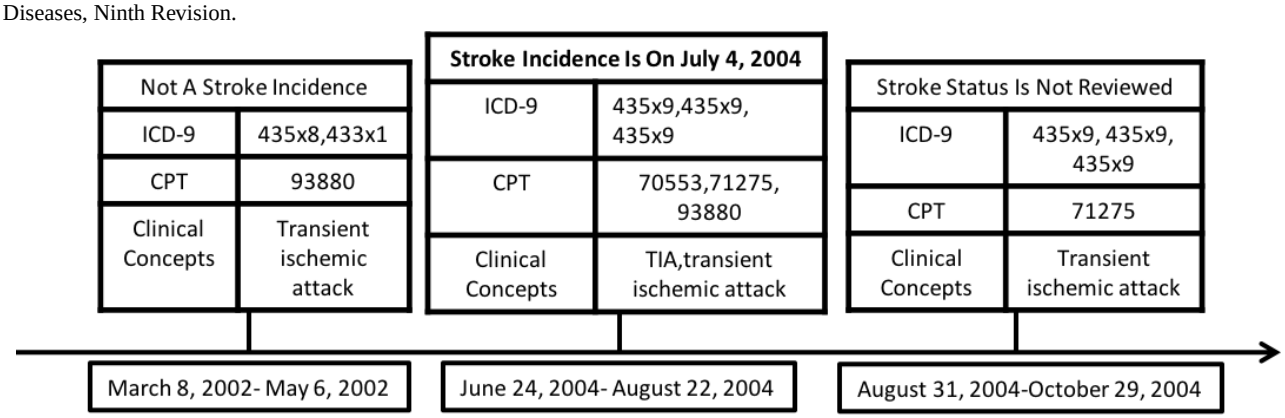
Candidate Predictive Features

The proposed algorithm aimed to identify first (incident) stroke events within a certain time frame. The three major data elements we used were clinical concepts, ICD-9 codes, and CPT codes. To align with the manual review process, only codes and clinical notes from the AF incident date to March 31, 2015, were retrieved and processed. In our analyses, we constructed different models by varying the inclusion of CPT codes and symptom-related clinical concepts in the model feature set and compared different models' performances.

Both ICD-9 and CPT codes were extracted from the REP database. Clinical concepts were identified from the major and secondary problem list section of Mayo Clinic EHR, and from clinical notes from other REP sites using a natural language processing system, MedTagger [47]. Expert-provided vocabulary was adopted from a previous study [48] to extract clinical concepts from unstructured clinical notes. MedTagger enables a series of natural language processing processes, including regular expression matching and positive, negative, or probable identification with ConText [49,50], and is insensitive to upper and lower case. MedTagger is also able to determine if the extracted clinical concepts are referring to the patients or their family members, or if the extracted clinical concepts are in present tense and thus are referring to a current event rather than a past medical condition. We considered only documents with positive, present-tense stroke mentions that were referring to patients themselves. Table S1 in Multimedia Appendix 1 lists clinical concepts for 2 major stroke subtypes and stroke-related symptoms. Table S2 in Multimedia Appendix 1 lists ICD-9 codes for 2 stroke subtypes and stroke-related symptoms. Table S3 in Multimedia Appendix 1 lists the CPT codes used in the stroke algorithm.

Clinical concept dates were determined by the date of the clinical notes from which clinical concepts were extracted, while ICD-9 and CPT code dates were extracted from the REP. Each visit was characterized by clinical concepts, ICD-9, and CPT codes within a 60-day window. The visit date was determined by the earliest date of any of the 3 elements in the 60-day window. If visit dates were within a 60-day window of a confirmed stroke incidence date, they were considered positive instances; otherwise, they were considered negative instances. Figure 1 demonstrates an example with an incident stroke on July 4, 2004. All visits were extracted and included in our data set if there was at least one key word or code during a 60-day window. Nurse abstractors reviewed every visit sequentially until they determined the incidence date to be July 4, 2004. All subsequent visits after a positive stroke incident were not reviewed and thus were not included in our analyses. Since the confirmed stroke incidence date fell in the date range of the third visit (June 24, 2004-August 22, 2004), we considered the combination of codes and clinical concepts in this visit to be predictive of a positive stroke incidence.

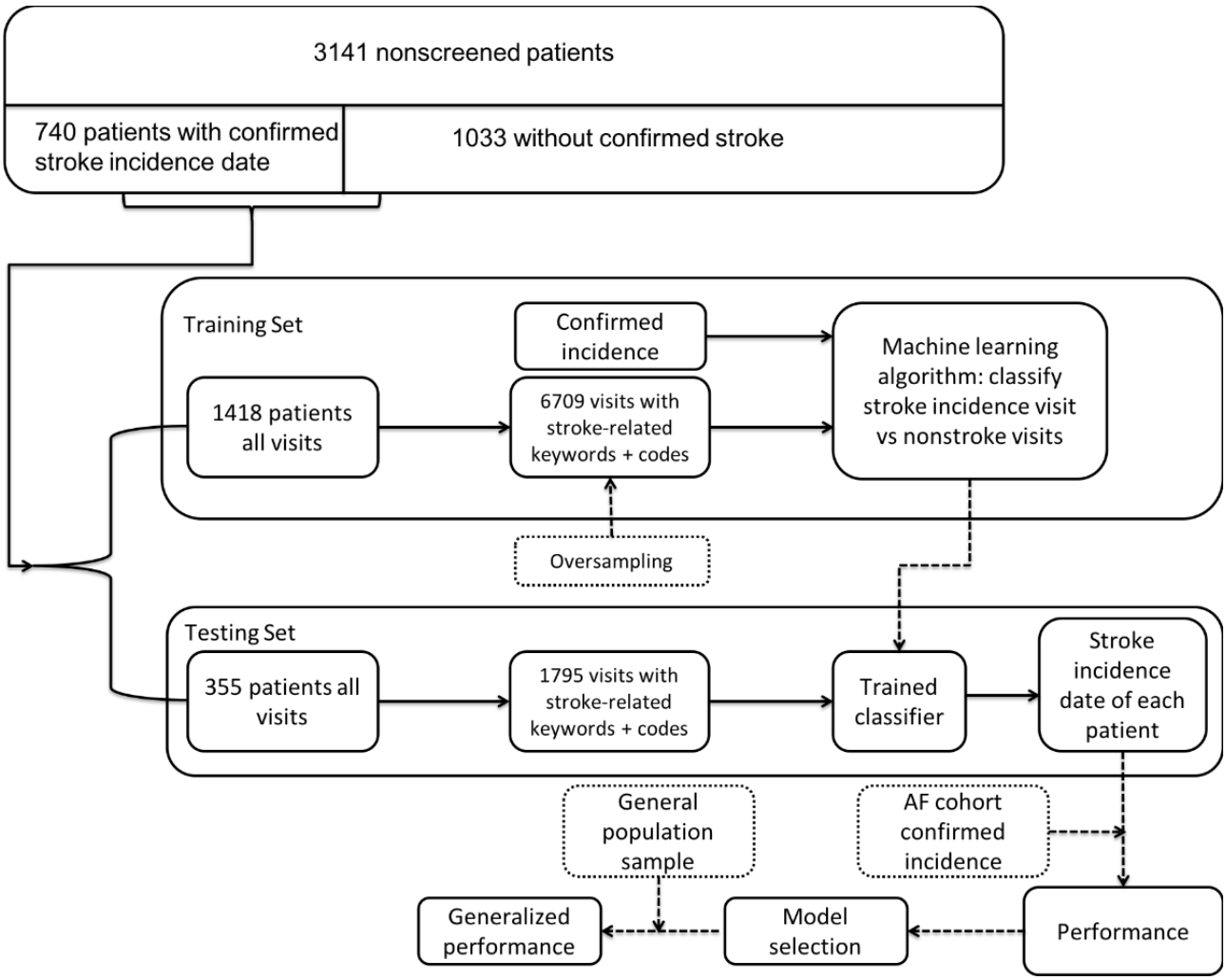
Figure 1. Inclusion of clinical concepts and codes on a patient visit timeline. CPT: Current Procedural Terminology; ICD-9: International Classification of Diseases, Ninth Revision.



Data Analysis

After incident stroke was confirmed, visits afterwards were not reviewed by abstractors and thus excluded from our overall data set. Figure 2 shows the workflow of the algorithm training and testing process. We created a data set with 9130 confirmed visits (with stroke vs nonstroke labels) among the 1773 patients. In total, there were 746 stroke visits and 8384 nonstroke visits. The stroke incidence count (n=746) was larger than the number of patients with confirmed stroke incidence (740) because incidence dates for different subtypes of stroke (ischemic stroke/transient ischemic attack and hemorrhagic stroke) were all recorded, such that patients might have had multiple incidence dates. We included data from a randomly selected 79.98% of our screened patients (1418/1773 patients; 7253 visits) as a training set and the remaining 20.02% of our screened patients (355/1773 patients; 1877 visits) were retained as an independent testing set. Due to the outcome imbalance in the data set (positive:negative ratio of about 1:10), we used the synthetic minority oversampling technique [51] to create oversampled training data sets with an oversampling percentage of 1000%.

Figure 2. Stroke algorithm training and testing workflow. AF: atrial fibrillation.



We considered two machine learning classifiers, logistic regression and random forest [52], to train our phenotyping models. Logistic regression served as a baseline modeling algorithm. Random forest was also chosen because of its high performance with structured input features and better model flexibility. We also considered the influence of feature groups by varying the inclusion of CPT codes and symptom terms in the input feature set. The hyperparameter tuning of the machine learning models was performed using 10-fold cross-validation. The performance metrics adopted for the machine learning task in the test set were precision, recall, and F score. The oversampling and machine learning modeling training and testing processes were implemented in Weka 3 (University of Waikato) [53]. Additional statistical summaries were performed using the R statistical software version 3.6.2 (The R Foundation for Statistical Computing). Quantitative variables are summarized as means, while nominal variables are expressed by counts and percentages.

Validation Cohort

We evaluated the generalizability of our model on a sample from a general population cohort of 71,429 patients. This cohort consisted of individuals sampled in Olmsted County, Minnesota on January 1, 2006, with an age ≥30 years and with no prior history of cardiovascular disease. We applied the best performing model based on the leave-out test set to this entire population cohort to generate incident stroke predictions. We then randomly selected 50 patients from those who had no stroke-related features (ie, de facto negative stroke predictions), 50 patients from those who were shown to have negative stroke

predictions, and 50 patients from those who were shown to have positive stroke predictions and a predicted incident stroke for evaluation. This verification-based sampling strategy allowed for estimates of positive and negative predictive values (PPVs and NPVs, respectively) by conditioning on algorithm predictions. Under these conditions (n=50), the half-width of the 95% Wilson score CI for the PPVs and separate NPVs would be approximately 0.1 for a true value of 0.85.

All 150 patient cases were reviewed by 1 nurse abstractor to confirm incident stroke, which served as our gold standard. We recorded model prediction outputs on all patient visits in the 150-patient validation set. We combined visit-level true predictions to generate patient-level incidence predictions by saving only the earliest date of positive predictions as stroke incidences. We compared patient-level incidence predictions with our gold standard. True prediction in our evaluation meant the date of the predicted incident stroke was within 60 days of the abstracted stroke date. A 2 x 2 confusion matrix was used to calculate performance scores for prediction evaluation. Model performance metrics included PPV and NPV using manual evaluation as the gold standard and patient-level predictions to calculate true positives, false positives, true negatives, and false negatives. The uncertainty of these performance estimates was calculated using Wilson score 95% CI for proportions.

In addition, we developed heuristic rules to distinguish stroke subtype (ischemic stroke/transient ischemic attack or hemorrhagic stroke) of each identified stroke incidence by analyzing the composition of keyword or code input feature sets (in a window of 60 days). We counted the number of keywords or codes for each ischemic stroke/transient ischemic attack and hemorrhagic stroke. If an input feature set contained more keywords or codes for ischemic stroke/transient ischemic attack, then this incidence was considered an ischemic stroke incidence; otherwise, it was considered a hemorrhagic stroke incidence. We only evaluated correct incident stroke predictions from the previous step in the evaluation data set with manually ascertained subtypes as the gold standard. Accuracy was calculated to measure performance of the subtype identification.

Results

Model Selection and Subtype Identification

Table 2 shows the algorithm performance measured on the test set for 8 models run on 4 input combinations and 2 classifiers (logistic regression and random forest). The random forest classifier outperformed the logistic classifier regardless of the feature sets used. Inclusion of CPT codes as features improved the performance for the random forest model with F score increased from 0.836 (Model 3) to 0.905 (Model 1). However, in the logistic model, the inclusion of CPT codes slightly improved the F score from 0.772 (Model 4) to 0.793 (Model 2). Using comparisons to all features (Model 1 and 2) and excluding the symptom terms (Model 6 and 7) achieved better F score (values italicized in Table 2).

Table 2. Stroke algorithm performance.

Model	ICD-9[a]	Clinical concept	CPT[b]	Classifier	Precision	Recall	F score
1	Yes	Symptoms + disease concepts	Yes	Random forest	0.912	0.906	0.905
2	Yes	Symptoms + disease concepts	Yes	Logistic	0.807	0.795	0.793
3	Yes	Symptoms + disease concepts	No	Random forest	0.835	0.845	0.836
4	Yes	Symptoms + disease concepts	No	Logistic	0.791	0.777	0.772
5	Yes	Disease-only concept	Yes	Random forest	*0.920*	*0.915*	*0.915*
6	Yes	Disease-only concept	Yes	Logistic	0.809	0.798	0.796
7	Yes	Disease-only concept	No	Random forest	0.856	0.847	0.846
8	Yes	Disease-only concept	No	Logistic	0.779	0.767	0.763

[a]ICD-9: International Classification of Diseases, Ninth Revision.

[b]CPT: Current Procedural Terminology.

Model Generalizability

Table 3 shows the distribution of stroke features in the AF cohort and the general population cohort. The AF cohort had a higher proportion of stroke-related codes and concepts. Results from the evaluation of the 150 selected patient records are presented in Table 4. Prediction performance corresponded to a PPV of 0.86 (95% CI 0.74-0.93), an NPV without ICD codes of 1.00 (95% CI 0.92-1.00), and an NPV with codes of 0.92 (95% CI 0.90-0.98). No strokes were observed among patients with no eligible stroke ICD codes. For subtype characterization, we achieved an accuracy of 80% (95% CI 0.68-0.89) in the general population sample.

Table 3. Patient feature distribution post-AF.

Stroke feature distribution	AF[a] screened		AF nonscreened (n=3141), n (%)	AF cohort total (n=4914), n (%)	Olmsted County cohort (N=71,429), n (%)
	Stroke (n=740), n (%)	No stroke (n=1033), n (%)			
ICD-9[b]+CPT[c]+CC[d]	654 (88.37)	379 (36.69)	0 (0)	1033 (21.02)	2726 (3.82)
ICD-9+CPT	66 (8.92)	596 (57.70)	0 (0)	662 (13.47)	1018 (1.42)
ICD-9+CC	9 (1.22)	12 (1.16)	0 (0)	21 (0.43)	48 (0.067)
CPT+CC	0 (0)	0 (0)	167 (5.32)	167 (3.40)	1595 (2.23)
ICD-9	11 (1.49)	46 (4.45)	0 (0)	57 (1.16)	194 (0.27)
CPT	0 (0)	0 (0)	1736 (55.27)	1736 (35.33)	17,433 (24.40)
CC	0 (0)	0 (0)	11 (0.35)	11 (0.24)	566 (0.79)
None	0 (0)	0 (0)	1227 (39.06)	1227 (24.97)	47,849 (66.99)

[a]AF: atrial fibrillation.

[b]ICD-9: International Classification of Diseases, Ninth Revision.

[c]CPT: Current Procedural Terminology.

[d]CC: clinical concepts.

Table 4. Generalizability analysis results from the Olmsted County cohort.

Gold standard	Stroke algorithm prediction (N=150)		
	Negative (n=100)		Positive (n=50)
	No ICD-9[a] codes (n=50)	Predicted no stroke (n=50)	
Stroke	0	4	43
No Stroke	50	46	7

[a]ICD-9: International Classification of Diseases, Ninth Revision.

Discussion

Principal Findings

The rapid expansion of information available in EHRs opens new opportunities to combine structured and unstructured data for research. Advances in machine learning methods and tools facilitate the combination of multimodal clinical data for effective development of phenotyping algorithms. However, performance of stroke electronic phenotyping algorithms varies by stroke subtypes [25] and phenotyping tasks (ie, case vs noncase or incident stroke phenotyping). Our previous study showed that when naïve ICD codes with clinical concept matching were used, stroke incidence identification had a PPV of 60.6% while case-versus-noncase identification had a much higher PPV of 88.7% [20].

In this study, we included clinical concepts extracted from clinical notes along with ICD-9 and CPT codes for incident stroke ascertainment. The rationale to add CPT codes is that diagnosis of stroke usually needs to be confirmed by imaging evidence and will probably be followed by therapeutic procedures. Thus, the addition of CPT codes in the model could potentially help to reduce the information redundancy effect by distinguishing between past and current events recorded in clinical notes. Our algorithm closely resembles the ascertainment process (chart review) of clinicians, which uses multiple types of EHR data (eg, diagnoses and procedure codes, unstructured clinical notes) in a parsimonious manner. Due to the redundancy and temporal ambiguity in unstructured clinical notes, we needed to construct a data set with sufficient and interpretable features from multimodal clinical data.

We found that the random forest generated better results, while the addition of CPT codes improved overall performance. This may be because imaging procedures, especially head CT or MRI, are critical in the diagnosis of stroke. Therefore, CPT codes of such procedures can be important indicators for distinguishing between incident and historical events. In addition, ICD codes and therapeutic procedures can vary significantly between incident and recurrent events. Meanwhile, we observed that the additions of stroke-related symptom concepts were not helpful for the phenotyping task. This may be due to the fact that our stroke incidence ascertainment depends largely on the ubiquitous nature of many stroke-related symptoms: they may be stroke-related but not necessarily stroke specific. Additionally, ascertainment requires well-documented evidence, such as imaging or imaging reports. Without properly recorded evidence, patients are not likely to be ascertained as stroke.

Our generalizability evaluation demonstrates that models trained using a specific disease cohort for incident stroke ascertainment can generalize well to a general patient population. This is very

encouraging given there are many existing patient cohorts available. Secondary use of these patient cohorts would be a cost-effective way for developing machine learning–based phenotyping algorithms. The study also illustrates that incorporating structured EHR data, such as CPT codes, can effectively distinguish incident stroke mentions from historical events in the clinical notes.

One limitation of our study is the dependence of domain experts to provide relevant clinical concepts, ICD-9 codes, and CPT codes. In the future, we will explore advance feature engineering approaches to identify those relevant concepts or codes automatically or semiautomatically. We are also aware that our imbalance cohort data and oversampling strategies might have introduced overfitting. Although our evaluation in the general population proved the performance of the algorithm, in the future, we can adopt a case–control matching strategy to deal with imbalanced data and mitigate the potential overfitting issue. In addition, new treatment strategies (mechanical thrombectomy) to treat stroke have been in the market in recent years, and thus the features used in our algorithm could have different weights for predictions of events in different temporal settings. A more precise strategy could consider using different features for prediction tasks in different time frames, where variations in clinical knowledge and care path have been considered.

Conclusions

In conclusion, the high prevalence of stroke and the lack of an efficient algorithm to confirm incident stroke events necessitate the development of an effective and interpretable algorithm to identify incident stroke occurrences. In this paper, we described our efforts to develop and validate an EHR-based algorithm that accurately identifies incident stroke events and goes beyond typical case-versus-noncase stroke identification. Our algorithm's good performance in a general population sample demonstrates its generalizability and potential to be adopted by other institutions.

Acknowledgments

The authors thank the other investigators, the staff, and the participants of the stroke phenotyping study for their valuable contributions. The stroke phenotyping study was conducted by a collaborative team of researchers from the Department of Health Sciences Research at Mayo Clinic and used 2 cohorts: the AF cohort and the Olmsted County cohort. This study was supported by grants from the National Institutes of Health (no. R01 HL136659 and R21 AG062580) and the American Heart Association (no. 11SDG7260039), and was made possible using the resources of the Rochester Epidemiology Project (no. R01 AG034676).

Authors' Contributions

YZ had full access to all study data and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors were responsible for study concept; design, acquisition, analysis, or interpretation of data; critical revision of the manuscript for important intellectual content; and administrative, technical, or material support. YZ was responsible for the drafting of the manuscript. YZ and NBL were responsible for statistical analysis. SB and AMC were responsible for obtaining funding. HL, SB, and NBL were responsible for study supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
ICD-9, CPT codes, and clinical concepts for two major stroke subtypes and stroke-related symptoms.
[DOCX File , 34 KB - jmir_v23i3e22951_app1.docx]

References

1. Bonita R. Epidemiology of stroke. Lancet 1992 Mar 08;339(8789):342-344. [doi: 10.1016/0140-6736(92)91658-u] [Medline: 1346420]
2. Clinical diagnosis of stroke subtypes. UptoDate. URL: https://www.uptodate.com/contents/clinical-diagnosis-of-stroke-subtypes [accessed 2020-07-22]
3. Morley KI, Wallace J, Denaxas SC, Hunter RJ, Patel RS, Perel P, et al. Defining disease phenotypes using national linked electronic health records: a case study of atrial fibrillation. PLoS One 2014;9(11):e110900 [FREE Full text] [doi: 10.1371/journal.pone.0110900] [Medline: 25369203]
4. Chamberlain AM, Brown RD, Alonso A, Gersh BJ, Killian JM, Weston SA, et al. No decline in the risk of stroke following incident atrial fibrillation since 2000 in the community: a concerning trend. J Am Heart Assoc 2016 Jun 13;5(6):e003408 [FREE Full text] [doi: 10.1161/JAHA.116.003408] [Medline: 27412902]
5. Scannapieco FA, Bush RB, Paju S. Associations between periodontal disease and risk for atherosclerosis, cardiovascular disease, and stroke. A systematic review. Ann Periodontol 2003 Dec;8(1):38-53. [doi: 10.1902/annals.2003.8.1.38] [Medline: 14971247]
6. Slee VN. The International Classification of Diseases: ninth revision (ICD-9). Ann Intern Med 1978 Mar;88(3):424-426. [doi: 10.7326/0003-4819-88-3-424] [Medline: 629506]

7. Imran TF, Posner D, Honerlaw J, Vassy JL, Song RJ, Ho YL, et al. A phenotyping algorithm to identify acute ischemic stroke accurately from a national biobank: the Million Veteran Program. Clin Epidemiol 2018;10:1509-1521 [FREE Full text] [doi: 10.2147/CLEP.S160764] [Medline: 30425582]
8. Coull A, Silver L, Bull L, Giles M, Rothwell P, Oxford Vascular (OXVASC) Study. Direct assessment of completeness of ascertainment in a stroke incidence study. Stroke 2004 Sep;35(9):2041-2045. [doi: 10.1161/01.STR.0000137605.48864.2f] [Medline: 15256682]
9. Thangaraj P, Kummer B, Lorberbaum T, Elkind M, Tatonetti N. Comparative analysis, applications, and interpretation of electronic health record-based stroke phenotyping methods. BioData Min 2020 Dec 07;13(1):21 [FREE Full text] [doi: 10.1186/s13040-020-00230-x] [Medline: 33372632]
10. Moons KGM, Bots ML, Salonen JT, Elwood PC, Freire de Concalves A, Nikitin Y, et al. Prediction of stroke in the general population in Europe (EUROSTROKE): Is there a role for fibrinogen and electrocardiography? J Epidemiol Community Health 2002 Mar;56 Suppl 1:i30-i36 [FREE Full text] [doi: 10.1136/jech.56.suppl_1.i30] [Medline: 11815642]
11. Ho KC, Speier W, El-Saden S, Liebeskind DS, Saver JL, Bui AAT, et al. Predicting discharge mortality after acute ischemic stroke using balanced data. AMIA Annu Symp Proc 2014;2014:1787-1796 [FREE Full text] [Medline: 25954451]
12. Peng S, Chuang Y, Kang T, Tseng K. Random forest can predict 30-day mortality of spontaneous intracerebral hemorrhage with remarkable discrimination. Eur J Neurol 2010 Jul;17(7):945-950. [doi: 10.1111/j.1468-1331.2010.02955.x] [Medline: 20136650]
13. Colak C, Karaman E, Turtay MG. Application of knowledge discovery process on the prediction of stroke. Comput Methods Programs Biomed 2015 May;119(3):181-185. [doi: 10.1016/j.cmpb.2015.03.002] [Medline: 25827533]
14. Asadi H, Dowling R, Yan B, Mitchell P. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. PLoS One 2014;9(2):e88225 [FREE Full text] [doi: 10.1371/journal.pone.0088225] [Medline: 24520356]
15. Denny JC, Crawford DC, Ritchie MD, Bielinski SJ, Basford MA, Bradford Y, et al. Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. Am J Hum Genet 2011 Oct 07;89(4):529-542 [FREE Full text] [doi: 10.1016/j.ajhg.2011.09.008] [Medline: 21981779]
16. Conway M, Berg R, Carrell D, Denny J, Kho A, Kullo I, et al. Analyzing the heterogeneity and complexity of Electronic Health Record oriented phenotyping algorithms. AMIA Annu Symp Proc 2011;2011:274-283 [FREE Full text] [Medline: 22195079]
17. Overby CL, Pathak J, Gottesman O, Haerian K, Perotte A, Murphy S, et al. A collaborative approach to developing an electronic health record phenotyping algorithm for drug-induced liver injury. J Am Med Inform Assoc 2013 Dec;20(e2):e243-e252 [FREE Full text] [doi: 10.1136/amiajnl-2013-001930] [Medline: 23837993]
18. Castro VM, Minnier J, Murphy SN, Kohane I, Churchill SE, Gainer V, International Cohort Collection for Bipolar Disorder Consortium. Validation of electronic health record phenotyping of bipolar disorder cases and controls. Am J Psychiatry 2015 Apr;172(4):363-372 [FREE Full text] [doi: 10.1176/appi.ajp.2014.14030423] [Medline: 25827034]
19. Feigin VL, Carter K. Editorial comment--Stroke incidence studies one step closer to the elusive gold standard? Stroke 2004 Sep;35(9):2045-2047. [Medline: 15331801]
20. Hollands GJ, Marteau TM, Fletcher PC. Non-conscious processes in changing health-related behaviour: a conceptual analysis and framework. Health Psychology Review 2016 Feb 16;10(4):381-394 [FREE Full text] [doi: 10.1080/17437199.2015.1138093]
21. Ni Y, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, et al. Towards phenotyping stroke: leveraging data from a large-scale epidemiological study to detect stroke diagnosis. PLoS One 2018;13(2):e0192586 [FREE Full text] [doi: 10.1371/journal.pone.0192586] [Medline: 29444182]
22. Howard VJ, McClure LA, Meschia JF, Pulley L, Orr SC, Friday GH. High prevalence of stroke symptoms among persons without a diagnosis of stroke or transient ischemic attack in a general population: the REasons for Geographic And Racial Differences in Stroke (REGARDS) study. Arch Intern Med 2006 Oct 09;166(18):1952-1958. [doi: 10.1001/archinte.166.18.1952] [Medline: 17030827]
23. How a CPT code becomes a code. American Speech-Language-Hearing Association. URL: https://www.asha.org/uploadedFiles/How-A-CPT-Code-Becomes-A-Code.pdf [accessed 2020-07-22]
24. Denaxas SC, George J, Herrett E, Shah AD, Kalra D, Hingorani AD, et al. Data resource profile: cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). Int J Epidemiol 2012 Dec;41(6):1625-1638 [FREE Full text] [doi: 10.1093/ije/dys188] [Medline: 23220717]
25. Woodfield R, Grant I, UK Biobank Stroke Outcomes Group, UK Biobank Follow-Up and Outcomes Working Group, Sudlow CLM. Accuracy of electronic health record data for identifying stroke cases in large-scale epidemiological studies: a systematic review from the UK Biobank Stroke Outcomes Group. PLoS One 2015;10(10):e0140533 [FREE Full text] [doi: 10.1371/journal.pone.0140533] [Medline: 26496350]
26. Schuemie MJ, Sen E, 't JGW, van SEM, Sturkenboom MC, Kors JA. Automating classification of free-text electronic health records for epidemiological studies. Pharmacoepidemiol Drug Saf 2012 Jun;21(6):651-658. [doi: 10.1002/pds.3205] [Medline: 22271492]

27. Schellenbaum GD, Heckbert SR, Smith NL, Rea TD, Lumley T, Kitzman DW, et al. Congestive heart failure incidence and prognosis: case identification using central adjudication versus hospital discharge diagnoses. Ann Epidemiol 2006 Mar;16(2):115-122. [doi: 10.1016/j.annepidem.2005.02.012] [Medline: 15964203]
28. Pakhomov S, Weston SA, Jacobsen SJ, Chute CG, Meverden R, Roger VL. Electronic medical records for clinical research: application to the identification of heart failure. Am J Manag Care 2007 Jun;13(6 Part 1):281-288 [FREE Full text] [Medline: 17567225]
29. Ermenc B. Minimizing mistakes in clinical diagnosis. J Forensic Sci 1999 Jul;44(4):810-813. [Medline: 10432615]
30. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, eMERGE Team. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC Med Genomics 2011 Jan 26;4:13 [FREE Full text] [doi: 10.1186/1755-8794-4-13] [Medline: 21269473]
31. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med 2013 Oct;15(10):761-771 [FREE Full text] [doi: 10.1038/gim.2013.72] [Medline: 23743551]
32. Denny JC. Chapter 13: Mining electronic health records in the genomics era. PLoS Comput Biol 2012;8(12):e1002823 [FREE Full text] [doi: 10.1371/journal.pcbi.1002823] [Medline: 23300414]
33. Kho AN, Hayes MG, Rasmussen-Torvik L, Pacheco JA, Thompson WK, Armstrong LL, et al. Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. J Am Med Inform Assoc 2012;19(2):212-218 [FREE Full text] [doi: 10.1136/amiajnl-2011-000439] [Medline: 22101970]
34. Ritchie MD, Denny JC, Zuvich RL, Crawford DC, Schildcrout JS, Bastarache L, Cohorts for HeartAging Research in Genomic Epidemiology (CHARGE) QRS Group, et al. Genome- and phenome-wide analyses of cardiac conduction identifies markers of arrhythmia risk. Circulation 2013 Apr 02;127(13):1377-1385 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.112.000604] [Medline: 23463857]
35. Wright A, Pang J, Feblowitz JC, Maloney FL, Wilcox AR, Ramelson HZ, et al. A method and knowledge base for automated inference of patient problems from structured data in an electronic medical record. J Am Med Inform Assoc 2011;18(6):859-867 [FREE Full text] [doi: 10.1136/amiajnl-2011-000121] [Medline: 21613643]
36. Garvin JH, DuVall SL, South BR, Bray BE, Bolton D, Heavirland J, et al. Automated extraction of ejection fraction for quality measurement using regular expressions in Unstructured Information Management Architecture (UIMA) for heart failure. J Am Med Inform Assoc 2012;19(5):859-866 [FREE Full text] [doi: 10.1136/amiajnl-2011-000535] [Medline: 22437073]
37. Percha B, Nassif H, Lipson J, Burnside E, Rubin D. Automatic classification of mammography reports by BI-RADS breast tissue composition class. J Am Med Inform Assoc 2012;19(5):913-916 [FREE Full text] [doi: 10.1136/amiajnl-2011-000607] [Medline: 22291166]
38. Jiang M, Chen Y, Liu M, Rosenbloom ST, Mani S, Denny JC, et al. A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. J Am Med Inform Assoc 2011;18(5):601-606 [FREE Full text] [doi: 10.1136/amiajnl-2011-000163] [Medline: 21508414]
39. Xu Y, Hong K, Tsujii J, Chang EI. Feature engineering combined with machine learning and rule-based methods for structured information extraction from narrative clinical discharge summaries. J Am Med Inform Assoc 2012;19(5):824-832 [FREE Full text] [doi: 10.1136/amiajnl-2011-000776] [Medline: 22586067]
40. D'Avolio LW, Nguyen TM, Goryachev S, Fiore LD. Automated concept-level information extraction to reduce the need for custom software and rules development. J Am Med Inform Assoc 2011;18(5):607-613 [FREE Full text] [doi: 10.1136/amiajnl-2011-000183] [Medline: 21697292]
41. Kahn MG, Weng C. Clinical research informatics: a conceptual perspective. J Am Med Inform Assoc 2012 Jun;19(e1):e36-e42 [FREE Full text] [doi: 10.1136/amiajnl-2012-000968] [Medline: 22523344]
42. Chamberlain AM, Gersh BJ, Alonso A, Chen LY, Berardi C, Manemann SM, et al. Decade-long trends in atrial fibrillation incidence and survival: a community study. Am J Med 2015 Mar;128(3):260-7.e1 [FREE Full text] [doi: 10.1016/j.amjmed.2014.10.030] [Medline: 25446299]
43. St Sauver JL, Grossardt BR, Yawn BP, Melton LJ, Rocca WA. Use of a medical records linkage system to enumerate a dynamic population over time: the Rochester epidemiology project. Am J Epidemiol 2011 May 01;173(9):1059-1068 [FREE Full text] [doi: 10.1093/aje/kwq482] [Medline: 21430193]
44. Rocca WA, Yawn BP, St Sauver JL, Grossardt BR, Melton LJ. History of the Rochester Epidemiology Project: half a century of medical records linkage in a US population. Mayo Clin Proc 2012 Dec;87(12):1202-1213 [FREE Full text] [doi: 10.1016/j.mayocp.2012.08.012] [Medline: 23199802]
45. Rocca WA, Grossardt BR, Brue SM, Bock-Goodner CM, Chamberlain AM, Wilson PM, et al. Data Resource Profile: Expansion of the Rochester Epidemiology Project medical records-linkage system (E-REP). Int J Epidemiol 2018 Apr 01;47(2):368-368j [FREE Full text] [doi: 10.1093/ije/dyx268] [Medline: 29346555]
46. Witt BJ, Brown RD, Jacobsen SJ, Weston SA, Yawn BP, Roger VL. A community-based study of stroke incidence after myocardial infarction. Ann Intern Med 2005 Dec 06;143(11):785-792. [doi: 10.7326/0003-4819-143-11-200512060-00006] [Medline: 16330789]

47. Liu H, Bielinski SJ, Sohn S, Murphy S, Wagholikar KB, Jonnalagadda SR, et al. An information extraction framework for cohort identification using electronic health records. AMIA Jt Summits Transl Sci Proc 2013;2013:149-153 [FREE Full text] [Medline: 24303255]
48. Mitchell-Box K, Braun KL. Fathers' thoughts on breastfeeding and implications for a theory-based intervention. J Obstet Gynecol Neonatal Nurs 2012;41(6):E41-E50. [doi: 10.1111/j.1552-6909.2012.01399.x] [Medline: 22861175]
49. Harkema H, Dowling JN, Thornblade T, Chapman WW. ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports. J Biomed Inform 2009 Oct;42(5):839-851 [FREE Full text] [doi: 10.1016/j.jbi.2009.05.002] [Medline: 19435614]
50. Chapman W, Chu D, Dowling J. editors. ConText: An algorithm for identifying contextual features from clinical text. 2007. URL: https://dl.acm.org/doi/10.5555/1572392.1572408 [accessed 2020-07-21]
51. Chawla N, Bowyer K, Hall L, Kegelmeyer W. SMOTE: Synthetic Minority Over-sampling Technique. JAIR 2002 Jun 01;16(6):321-357. [doi: 10.1613/jair.953]
52. Liaw A, Weiner M. Classification and regression by randomForest. R news 2002;2(3):18-22.
53. Eibe F, Hall M, Witten I. The WEKA Workbench. In: Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann. Burlington, Massachusetts: Morgan Kaufmann; 4th edition (December 1, 2016); 2016.

Abbreviations

AF: atrial fibrillation
CPT: Current Procedural Terminology
CT: computed tomography
EHR: electronic health record
eMERGE: Electronic Medical Records and Genomics
ICD: International Classification of Diseases
ICD-9: International Classification of Diseases, Ninth Revision
MRI: magnetic resonance imaging
NPV: negative predictive value
PPV: positive predictive value
REP: Rochester Epidemiology Project

Edited by G Eysenbach, R Kukafka; submitted 27.07.20; peer-reviewed by C Colak, MO Kaya; comments to author 12.08.20; revised version received 25.08.20; accepted 20.01.21; published 08.03.21.

Please cite as:
Zhao Y, Fu S, Bielinski SJ, Decker PA, Chamberlain AM, Roger VL, Liu H, Larson NB
Natural Language Processing and Machine Learning for Identifying Incident Stroke From Electronic Health Records: Algorithm Development and Validation
J Med Internet Res 2021;23(3):e22951
URL: https://www.jmir.org/2021/3/e22951
doi:10.2196/22951
PMID:33683212

Original Paper

Machine Learning for Mental Health in Social Media: Bibliometric Study

Jina Kim[1], BSc, MSc; Daeun Lee[1], BA; Eunil Park[1], BSc, MSc, PhD

Department of Applied Artificial Intelligence, Sungkyunkwan University, Seoul, Republic of Korea

Corresponding Author:
Eunil Park, BSc, MSc, PhD
Department of Applied Artificial Intelligence
Sungkyunkwan University
312 International Hall
Sungkyunkwan-ro 25-2
Seoul, 03063
Republic of Korea
Phone: 82 2 740 1864
Fax: 82 2 740 1856
Email: eunilpark@skku.edu

Abstract

Background: Social media platforms provide an easily accessible and time-saving communication approach for individuals with mental disorders compared to face-to-face meetings with medical providers. Recently, machine learning (ML)-based mental health exploration using large-scale social media data has attracted significant attention.

Objective: We aimed to provide a bibliometric analysis and discussion on research trends of ML for mental health in social media.

Methods: Publications addressing social media and ML in the field of mental health were retrieved from the Scopus and Web of Science databases. We analyzed the publication distribution to measure productivity on sources, countries, institutions, authors, and research subjects, and visualized the trends in this field using a keyword co-occurrence network. The research methodologies of previous studies with high citations are also thoroughly described.

Results: We obtained a total of 565 relevant papers published from 2015 to 2020. In the last 5 years, the number of publications has demonstrated continuous growth with *Lecture Notes in Computer Science* and *Journal of Medical Internet Research* as the two most productive sources based on Scopus and Web of Science records. In addition, notable methodological approaches with data resources presented in high-ranking publications were investigated.

Conclusions: The results of this study highlight continuous growth in this research area. Moreover, we retrieved three main discussion points from a comprehensive overview of highly cited publications that provide new in-depth directions for both researchers and practitioners.

(J Med Internet Res 2021;23(3):e24870) doi:10.2196/24870

KEYWORDS

bibliometric analysis; machine learning; mental health; social media

Introduction

Background

Artificial intelligence (AI) has permeated various daily sectors that are directly related to our lives [1,2]. With this trend, AI for health, which refers to applying AI to real-world health care, has become one of the most important social issues at present [3,4]. With privacy and security as the bedrock of AI-based health care, there have been many attempts to employ AI and its applications in health care services [5,6]. As a representative example, Rizwan Malik, a radiologist in the United Kingdom, adopted a unique AI-based chest X-ray system to reduce patient waiting time in the COVID-19 pandemic scenario [4]. Furthermore, Microsoft [2] has invested approximately US $20 million to aid the collaboration teams of health care professionals and data science/AI experts in COVID-19–related research.

Extensive efforts have been put forward to employ AI technologies in health care services in addressing issues related to physical health, involving several medical centers, researchers, and organizations, as well as for mental health as a rapidly growing social issues. Although mental health is a pervasive and comprehensive issue, its detection and exposure are challenging. The World Health Organization estimates that approximately 1 billion people worldwide have mental disorders [7]. Moreover, 264 million people have been globally affected by depression, a common mental disorder [8]. However, more than 75% of people in underdeveloped countries (ie, low-income countries) suffering from mental disorders do not receive any treatments [7]. Several scholars have also revealed that individuals who suffer from mental disorders tend to prefer sharing their personal information and seeking assistance to reduce their concerns through online channels rather than with medical providers such as counselors or therapists [9-11].

Considering this tendency, social media represents a supportive tool for these individuals [11], where users are allowed to generate content, share information, and communicate [12]. Many researchers have attempted to explore the large-scale user-generated content in social media by means of machine learning (ML), which is a robust data-engineering technique, to analyze hidden information and knowledge on mental health. Therefore, we provide a theoretical background of related studies in the following subsections.

Related Review Papers

Higgins et al [13] define a systematic review as

> *a study that is composed of a search for scientific publications related to various topics in accordance with systematic guidelines including the search queries, the scientific databases, and the assessment criteria.*

With this concept, several prior mental health studies have investigated how to utilize ML in social media datasets. For instance, Seabrook et al [14] examined a systematic approach to provide an overview of prior research that focused on depression and anxiety in social media contexts between 2005 and 2016 with 8 identified databases. To objectively evaluate and summarize the literature, each case was evaluated by three unique dimensions: how to include psychological/cognitive measures, how to use external measurements for mental health criteria, and how to collect user activities in social media. Subsequently, 70 cases were selected, examined, and reviewed for both the implications and future directions regarding the application of ML to mental health in social media.

In addition to a systematic review approach, a scoping review may be performed, which is defined as "a type of research synthesis that aims to map the literature on a particular topic or research area and provide an opportunity to identify key concepts" [15]. This implies that a scoping review provides a bird's eye view of key concepts in specific research areas, main sources, findings, and implications. For instance, Shatte et al [16] adopted a scoping review approach including 300 papers that focused on ML and big data applications in mental health, and concluded that the majority of these papers considered depression, schizophrenia, and Alzheimer disease as their main mental illnesses. Moreover, 89% of the papers analyzed utilized supervised learning approaches such as support vector machine (SVM), naïve Bayes, or decision trees to examine their selected illness.

Chancellor et al [17] also performed a thematic discourse analysis on 55 scientific papers with the goal of predicting mental health status in social media, and demonstrated that interdisciplinary researchers have different perspectives toward users' datasets; these perspectives were classified as "human-centered machine learning" (HCML). Based on these findings and the concept of HCML, Chancellor and De Choudhury [18] subsequently categorized a total of 75 papers with five discourses: disorder/patient, social media, scientific, data/ML, and person. Based on this categorization, a total of 75 cases in which mental health status was assessed using social media datasets within 41 conference/journal papers published from 2013 to 2018 retrieved through academic databases (eg, ACM Digital Library and Google Scholar) were reviewed with respect to data annotation methods, data collection/quality management, preprocessing procedures, feature selections, model selection, and verification.

As presented in numerous prior studies, ML and mental health in social media have gained exponential attention in both practical and academic fields. Thus, we aimed to perform a bibliometric analysis to provide an overview and recent trends of this field.

Related Bibliometric Analyses

Bibliometric analysis is an extensive and widely used approach "to shed light on the processes of written communication and of the nature and course of development of a discipline" [19]. A bibliometric analysis thus allows researchers to understand the trends of specific research areas with several primary publications, including collaboration relations [20,21], core research themes [20], and scientific techniques [22].

Several scholars have performed bibliometric analyses on AI/ML in health care areas, including public and mental health, as well as in areas of specific mental illnesses. For example, dos Santos et al [22] performed a bibliometric analysis of data mining and ML techniques applied to public health issues based on papers published between 2009 and 2018 retrieved from three academic databases: Web of Science (WoS), Scopus, and ScienceDirect.

In the case of depression specifically, Tran et al [23] used a bibliometric approach to examine AI applications presented in publications indexed in WoS, evaluated the productivity of AI research through statistical analyses, and performed an exploratory factor analysis on the contexts of paper abstracts to present the most relevant and popular research issues. Moreover, Wang et al [24] performed a bibliometric analysis of natural language processing in various medical research areas including papers retrieved from PubMed data engines published from 1999 to 2018.

With respect to social media, several bibliometric analyses have closely evaluated relevant publications and their effects on society. As a representative example, Chen et al [20] adopted

both quantitative and statistical approaches with the WoS database to detect specific events in social media within the period of 2009-2017, investigated the number of publications and degree of collaboration, and further used clustering analysis to identify the main research themes. Another bibliometric study conducted by Sa'ed et al [21] focused on social media in psychology over 12 years based on records retrieved from WoS, and identified bibliometric indicators, including international collaboration/research networks.

Based on this background, the use of ML in medical fields and social media has been extensively explored using bibliometric approaches with notable implications and future directions. Therefore, this is an appropriate time to provide more detailed observations on ML with respect to the relation of specific medical areas with social media. Specifically, we examined the trends of research using ML for mental health in social media by employing (1) a bibliometric analysis to determine the publication distributions on sources (journals or conferences), authors, institutions, countries, research subjects, and author keywords; and (2) a trend review analysis to determine the distributions of citation numbers, along with a comprehensive review of highly cited publications.

With these approaches, we aimed to identify overall research trends of this area in a quantitative manner, and to qualitatively identify the key methodologies used on diverse social media platforms. These findings can shed light on the recent trends in the field and highlight more detailed directions of future research areas.

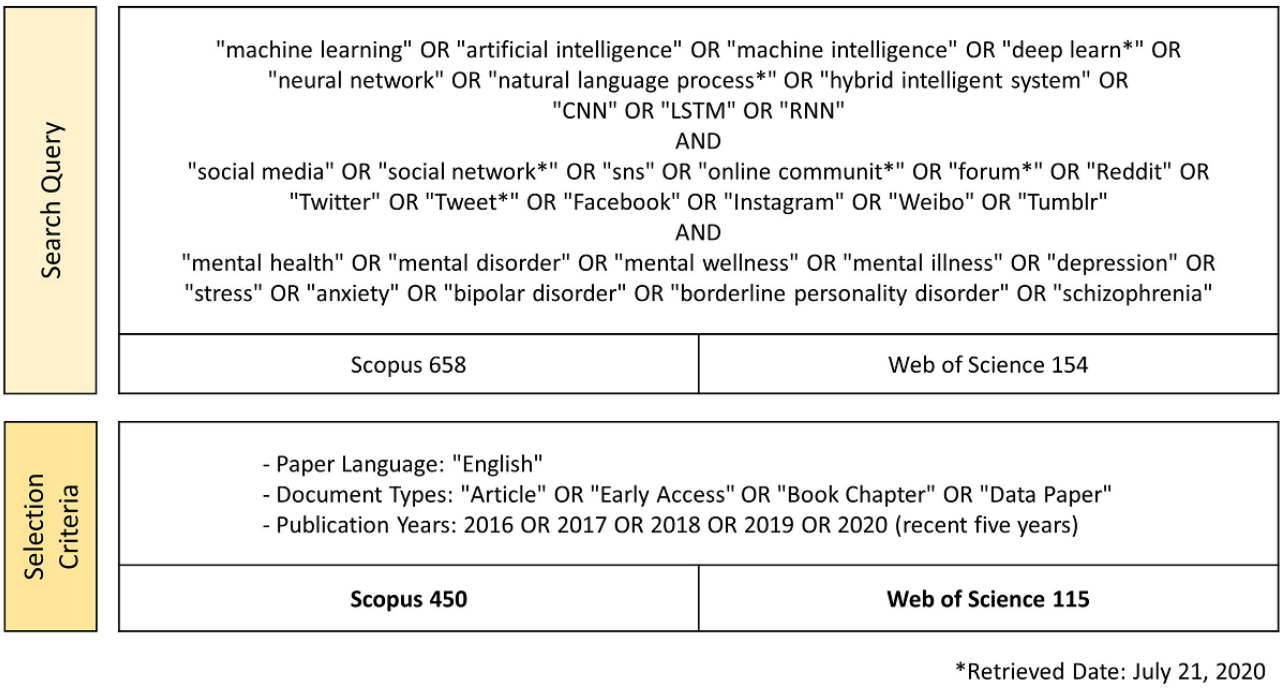
Methods

Data Collection

We collected papers from two citation databases, Scopus and WoS. Scopus is one of the largest citation repositories that covers scientific journals, conference proceedings, and books. WoS stores high-quality publications evaluated by three main indices: Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI), or Art & Humanities Citation Index (A&HCI).

Relevant publications were obtained when the terms included in the search query appeared in the title, abstract, or keywords. We defined the search query of each topic based on prior research on ML, social media [17], and mental health [25]. We excluded papers that were not written in English or were categorized as other document types (Figure 1). As a result, a total of 565 papers published from 2016 to 2020 were obtained on July 21, 2020. To cover rapidly changing trends in ML areas, we also considered the year 2020, which is still open for new issues. The complete list of included publications is provided in Multimedia Appendix 1.

Figure 1. Representative data collection procedure.



Analysis Methodologies

A bibliometric analysis includes the distribution exploration of publication and research subject, as well as citation quantities. Both the Python programming language and Microsoft Excel were employed to perform statistical analyses of the retrieved papers. We first analyzed the publication distributions of papers with several categories (eg, sources, countries, institutions, authors, and research subjects). We also performed a network analysis of frequently used keywords. Moreover, to identify the research trends in this area, we performed trend review analyses with highly cited papers covering the following topics: (i) ML techniques, (ii) specific mental illnesses, and (iii) social media.

Results

Publication Distribution Analysis

Overall Publication Trend

The continuous growth of publications from 2016 to 2020 (until July 2020) is illustrated in Table 1. In 2016, two papers were retrieved from WoS and 33 papers were retrieved from Scopus. The publication count demonstrates rapid growth in 2019 with 43 publications retrieved from WoS and 166 publications retrieved from Scopus. Considering the retrieved date (July 2020), we expect that more papers would be retrieved in the remainder of 2020 up to the present.

Table 1. Number of publications per year.

Year	Publication count, n (%)	
	Scopus (N=450)	Web of Science (N=115)
2016	33 (7.3)	2 (1.7)
2017	68 (15.1)	16 (13.9)
2018	88 (19.6)	21 (18.3)
2019	166 (36.9)	43 (37.4)
2020	95 (21.1)	33 (28.7)

Productive Publication Source

We considered several document types, including not only journal articles but also conference proceedings and book chapters. Tables 2 and 3 present the publication sources with high counts in Scopus and WoS, respectively. *Lecture Notes in Computer Science* was the most productive publication source in Scopus, followed by *CEUR Workshop Proceedings*, *Neural Computing and Applications*, and *Journal of Medical Internet Research* with more than 20 publication counts each. *Journal of Medical Internet Research* was selected as the most productive publication source in WoS with 15 publication counts, followed by *IEEE Access*.

Table 2. Top publication sources in Scopus (N=450).

Rank	Source	Publication count, n (%)
1	Lecture Notes in Computer Science	35 (7.8)
2	CEUR Workshop Proceedings	25 (5.6)
3	Neural Computing and Applications	22 (4.9)
4	Journal of Medical Internet Research	16 (3.6)
5	Advances in Intelligent Systems and Computing	13 (2.9)
6	ACM International Conference Proceeding Series	9 (2.0)
6	International Journal of Innovative Technology and Exploring Engineering	9 (2.0)
8	IEEE Access	7 (1.6)
9	Communications in Computer and Information Science	5 (1.1)
9	Frontiers in Psychiatry	5 (1.1)
9	International Journal of Environmental Research and Public Health	5 (1.1)

Table 3. Top publication sources in Web of Science (N=115).

Rank	Source	Publication count, n (%)
1	Journal of Medical Internet Research	15 (13.0)
2	IEEE Access	7 (6.1)
3	BMJ Open	3 (2.6)
3	Computers in Human Behavior	3 (2.6)
3	Frontiers in Psychology	3 (2.6)
3	IEEE Transactions on Knowledge and Data Engineering	3 (2.6)
3	International Journal of Environmental Research and Public Health	3 (2.6)
8	BMC Medical Informatics and Decision Making	2 (1.7)
8	Cyberpsychology Behavior and Social Networking	2 (1.7)
8	Journal of Information Science	2 (1.7)
8	Journal of Intelligent information Systems	2 (1.7)
8	Multimedia Tools and Applications	2 (1.7)
8	NPJ Schizophrenia	2 (1.7)
8	Scientific Reports	2 (1.7)
8	Social Science Computer Review	2 (1.7)
8	Translational Behavioral Medicine	2 (1.7)

Predominant Countries

More than 30 countries were identified as the predominant nations performing research in this field in Scopus (n=59) and WoS (n=39). Table 4 illustrates the top productive countries based on the number of publications. The United States was the most productive nation in both databases, followed by China and India.

Table 4. Top productive countries.

Rank	Country	Publication count, n (%)
Scopus (N=450)		
1	United States	146 (32.4)
2	India	66 (14.7)
3	China	63 (14.0)
4	United Kingdom	34 (7.6)
5	Canada	22 (4.9)
6	Spain	18 (4.0)
7	Australia	17 (3.8)
8	Germany	16 (3.6)
9	Taiwan	14 (3.1)
10	France	13 (2.9)
10	Netherlands	13 (2.9)
Web of Science (N=115)		
1	United States	52 (45.2)
2	China	25 (21.7)
3	United Kingdom	12 (10.4)
4	Australia	11 (9.6)
5	Spain	6 (5.2)
6	Canada	5 (4.4)
6	India	5 (4.4)
6	Saudi Arabia	5 (4.4)
7	South Korea	4 (3.5)
7	Taiwan	4 (3.5)

Productive Institutions

There were 391 different institutions associated with the 565 publications. The top-ranked institutions are presented in Table 5. Harvard University in the United States emerged as the most productive institution in WoS (13 publications), whereas Tsinghua University in China was selected as the most productive organization in Scopus (21 publications).

Table 5. Top productive institutions.

Institution	Publication count, n (%)	
	Scopus (N=450)	Web of Science (N=115)
Tsinghua University	21 (4.7)	3 (2.6)
Georgia Institute of Technology	12 (2.7)	4 (3.5)
Harvard University	14 (3.1)	13 (11.3)
University of Pennsylvania	13 (2.9)	5 (4.4)
Chinese Academy of Sciences	14 (3.1)	6 (5.2)
Johns Hopkins University	6 (1.3)	3 (2.6)
King's College London	6 (1.3)	—[a]
University of Toronto	6 (1.3)	3 (2.6)
National Tsing Hua University	6 (1.3)	3 (2.6)
Northwestern University	5 (1.1)	1 (0.9)
Centre National de la Recherche Scientifique	5 (1.1)	—
Vrije Universiteit Amsterdam	5 (1.1)	—
The University of Arizona	5 (1.1)	—
National University of Singapore	5 (1.1)	1 (0.9)
Deakin University	5 (1.1)	3 (2.6)
University of New South Wales	5 (1.1)	1 (0.9)
Ministry of Education China	4 (0.9)	—
Delhi Technological University	4 (0.9)	—
Cornell University	4 (0.9)	1 (0.9)
Radboud University Nijmegen	4 (0.9)	—
Russian Academy of Sciences	4 (0.9)	—
Microsoft Research	4 (0.9)	—
The University of Utah	4 (0.9)	1 (0.9)
Universidad Autónoma de Madrid	4 (0.9)	—
University of Rochester	4 (0.9)	—
University of Chinese Academy of Sciences	4 (0.9)	3 (2.6)
Université du Québec à Montréal	4 (0.9)	2 (1.7)
King Faisal University	4 (0.9)	2 (1.7)
University of Texas System	4 (0.9)	—
Asia University Taiwan	4 (0.9)	—

[a]—:no related records.

Predominant Authors

The top 20 researchers contributing to the field are listed in Table 6 based on their number of publications. Twelve researchers are affiliated to US-based organizations and five belong to Chinese institutions. The institutions of productive authors include not only several academic institutions but also some well-known hospitals such as Zucker Hillside Hospital. The most productive researcher was Professor Munmun De Choudhury, affiliated with Georgia Institute of Technology (15 publications), followed by Professor Sharath Chandra Guntuku from the University of Pennsylvania.

Table 6. Top 20 productive authors.

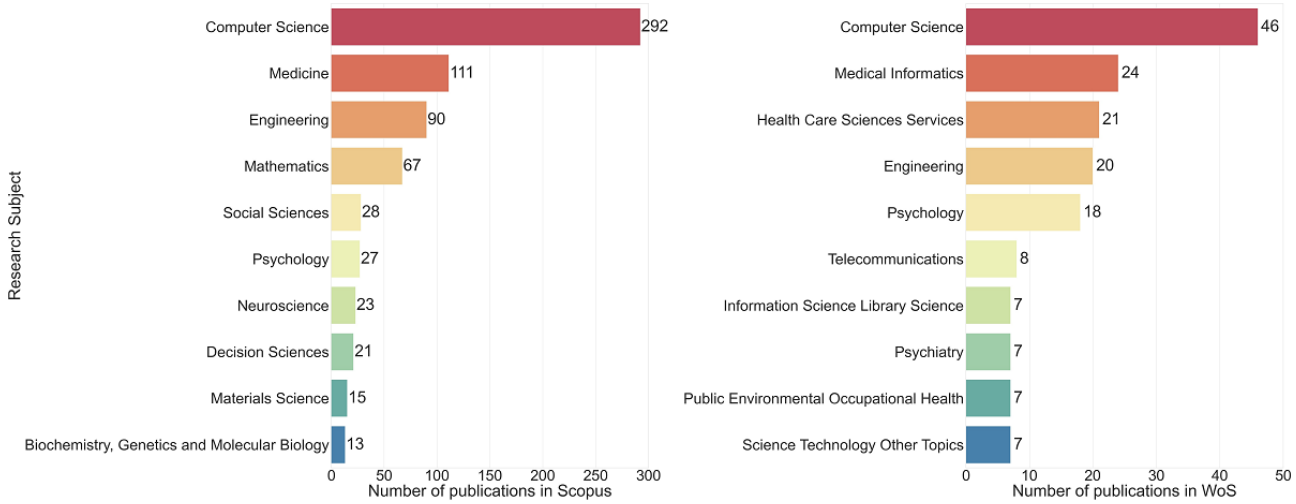
Author	Institution	Country	Publication count, n (%)	
			Scopus (N=450)	Web of Science (N=115)
M De Choudhury	Georgia Institute of Technology	United States	11 (2.4)	4 (3.5)
SC Guntuku	University of Pennsylvania	United States	5 (1.1)	4 (3.5)
HF Ahmad	King Faisal University	Saudi Arabia	4 (0.9)	3 (2.6)
SK Ernala	Georgia Institute of Technology	United States	4 (0.9)	3 (2.6)
LH Ungar	University of Pennsylvania	United States	4 (0.9)	3 (2.6)
T Nguyen	University of Pennsylvania	United States	5 (1.1)	2 (1.7)
S Venkatesh	University of Maryland	United States	5 (1.1)	2 (1.7)
ML Birnbaum	Zucker Hillside Hospital	United States	4 (0.9)	2 (1.7)
M Conway	University of Utah	United States	4 (0.9)	2 (1.7)
L Feng	Tsinghua University	China	4 (0.9)	2 (1.7)
D Phung	Deakin University	Australia	4 (0.9)	2 (1.7)
J Jia	Tsinghua University	China	6 (1.3)	—[a]
T Zhu	Chinese Academy of Sciences	China	6 (1.3)	—
RM Merchant	University of Pennsylvania	United States	2 (0.4)	3 (2.6)
H Christensen	University of New South Wales	Australia	3 (0.7)	2 (1.7)
JM Kane	Zucker Hillside Hospital	United States	3 (0.7)	2 (1.7)
Q Li	Tsinghua University	China	3 (0.7)	2 (1.7)
AF Rizvi	Zucker Hillside Hospital	United States	3 (0.7)	2 (1.7)
CY Shen	National Tsing Hua University	China	3 (0.7)	2 (1.7)
L Ungar	University of Pennsylvania	United States	3 (0.7)	2 (1.7)

[a]—: no related records.

Productive Research Subjects

The top 10 research subjects of each citation database are given in Figure 2. Among them, computer science was the most pivotal research subject in both databases (46, 40% in WoS; 292, 65% in Scopus). In Scopus, medicine (111, 25%), and engineering and mathematics (90, 20%) accounted for more than 10% of the total publications. In WoS, medical informatics (24, 21%), health care sciences services, engineering, and psychology constituted more than 10% of total publications.

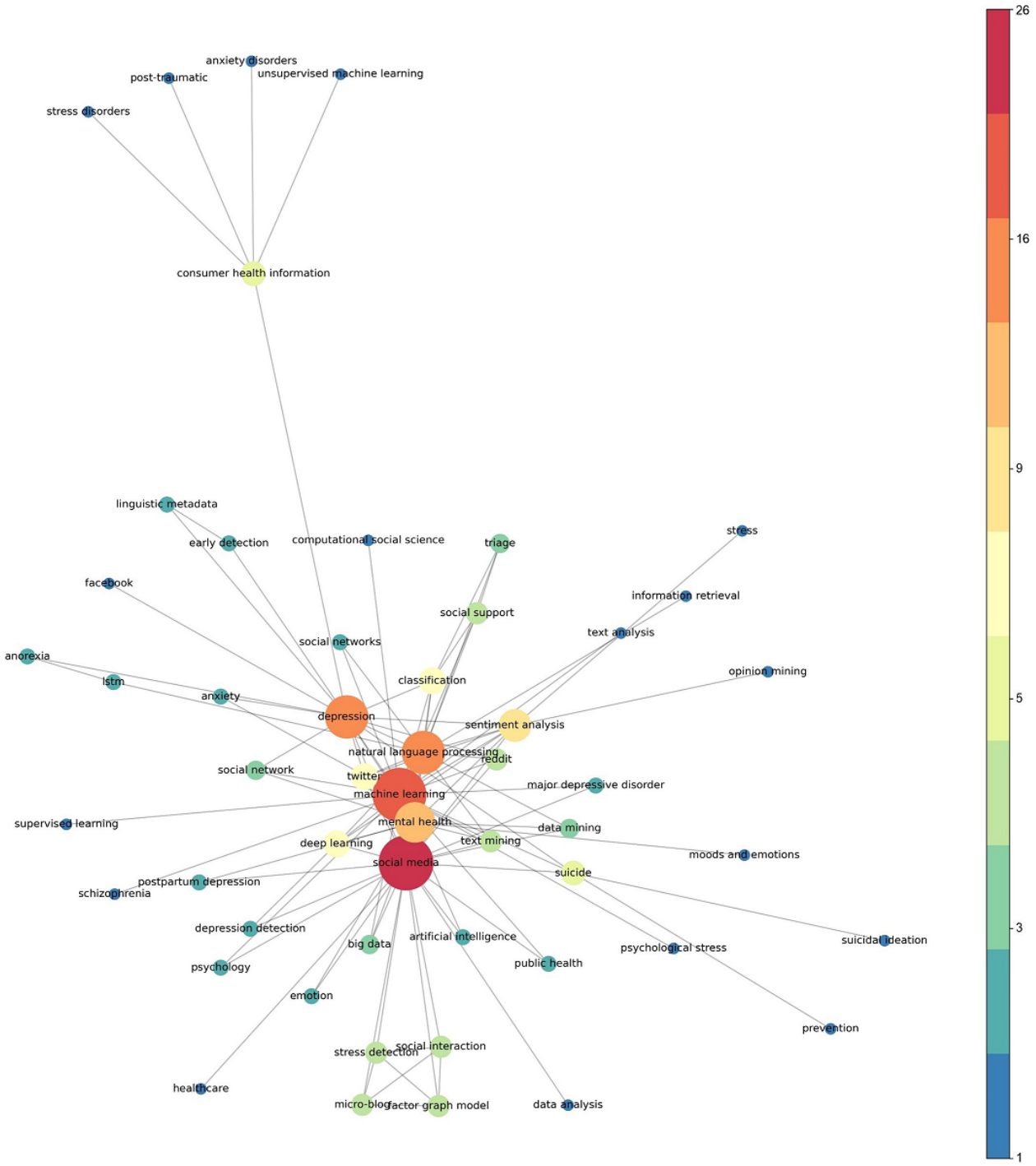
Figure 2. Publication count of top 10 research subjects.



Author Keyword Co-occurrence

Along with research subjects, we investigated the keywords derived by the authors, which represent the main research topics of the publications [26]. The keyword co-occurrence is visualized using a network graph (Figure 3), which is a well-known bibliometric methodology, in which each node is a keyword, while an edge between two nodes indicates the co-occurrence of two words. After building a network graph, we excluded edges with less than 3 co-occurrences.

Figure 3. Keyword co-occurrence network graph; the color map on the right side represents the degree centrality.



Note that the keywords with a high frequency reflect the scope of the research area, which includes social media, ML, natural language processing, and mental health. In particular, in the case of mental health–related words, depression was the most frequently presented keyword, followed by suicide, consumer health information, social support, and stress detection. The main research methodologies of papers included natural language processing, sentiment analysis, classification, text mining, and stress detection in the ML field. Twitter and Reddit were identified as the most widely investigated social media platforms in this area.

As there are several approaches of text analysis using natural language processing techniques in this field, we believe that future studies may provide hybrid approaches using both textual- and visual-based data collected from several types of social

media data. In addition, validating ML models trained on mental health–related social media data in clinical settings needs to be further investigated.

Overview of Highly Cited Publications

Publication Citation Quantities

The annual number of citations is presented in Table 7. Along with publication distributions, the annual number of citations has been consistently increasing. Up to July 21, 2020, more than 900 and 400 annual citations were recorded in Scopus and WoS, respectively.

Table 7. Number of citations per year.

Year	Citation count, n	
	Scopus	Web of Science
2016	14	1
2017	112	32
2018	349	120
2019	1000	341
2020	938	420

Comprehensive Analysis of Highly Cited Papers

We evaluated the overall academic output through bibliometric analysis. Due to the lack of observations in the content of the publications, as mentioned in previous studies [20,27], we extensively observed and reviewed the top five most highly cited papers per year to identify the comprehensive research methodologies in the field (Table 8). After excluding 7 duplicated papers, 39 papers were selected. Subsequently, a two-round filtering procedure was performed to determine whether a specific paper meets the following criteria: (i) addressing specific mental illness, (ii) using an ML technique, and (iii) utilizing datasets of social media. There were 15 papers that met these criteria. Subsequently, three experts in ML, medical services, and computer science, respectively, participated in the second-round filtering procedure. Following this, 10 papers that satisfied these criteria were selected.

Table 8. Overview of the top 5 most cited papers by year.

Year	Top-cited papers (N=39), n		Reviewed papers (N=10), n	
	Article	Conference paper	Article	Conference paper
2016	5	2	2	1
2017	6	0	3	0
2018	9	0	2	0
2019	8	2	1	0
2020	6	1	1	0

An overview and the research methodologies of the 10 highly cited publications are listed in Table 9, which are categorized according to the data source: Twitter (n=4 publications), Instagram (n=3), Facebook (n=3), Reddit (n=3), Weibo (n=2), and other online community (n=1).

Table 9. Summary of research methodologies employed in highly cited publications.

Data Source	Reference	Mental Health	Data Description	Machine learning				
				Model	Feature	Output	Annotation	Results
Twitter	Budenz et al [28]	Mental illness, bipolar disorder	1,270,902 tweets including bipolar or mental health-related words	Logistic regression	Term frequency-inverse document frequency	Related to mental illness or bipolar disorder	Manually annotated 2047 tweets with topic, stigma, and social support messaging	10-fold cross validation (AUC[a]=0.83)
Twitter	Du et al [29]	Suicide	1,962,766 tweets including 21 suicide-related keywords/phrases	CNN[b], SVM[c], extra trees, random forest, logistic regression, Bi-LSTM[d]	One-hot-vector mapped to pretrained GloVe Twitter embedding	Related to suicide	Manually annotated 3263 tweets and trained classifier to select 3000 additional suicide-related tweets	Accuracy 0.74, recall 0.96, precision 0.78, F1 0.83
Facebook, Twitter	Guntuku et al [30]	Psychological stress	601 users' Facebook and Twitter posts	Linear regression with several regularization methods (eg, ridge, elastic-net, LASSO[e] and L2 penalized SVMs)	LIWC[f], latent Dirichlet allocation topic modeling, stress lexicon, user engagement	Stress	Qualtrics survey; fill out the demographic questions and Cohen 10-item Stress Scale	5-fold cross-validation (Pearson r=0.24); trained on Facebook and Twitter, tested on Twitter
Facebook, Instagram	Shuai et al [31]	Social network mental disorders (eg, cyber-relationship addiction, information overload, and net compulsion)	3126 users' Instagram and Facebook data	Decision tree learning, SVM, logistic regression, DTSVM[g], SNMDD[h] (newly proposed model; tensor technique for deriving latent features)	Social interaction, personal profile, duration	Social network mental disorders	MTurk survey - fill out the standard social network mental disorder questionnaires; professional psychiatrists labeled users having a social network mental disorder	5-fold cross validation (accuracy0.78 for Instagram and 0.83 for Facebook)
Instagram	Reece and Danforth [32]	Depression	43,950 users' Instagram images	Random forest classifier	Number of comments and likes, number of faces in photograph, 3 color properties (hue, saturation, value)	Depression	MTurk survey; Center for Epidemiologic Studies Depression Scale to measure depression level	Recall: 0.697; precision: 0.604; F1: 0.647
Weibo	Lin et al [33]	Stress	1 billion Weibo posts	SVM, softmax regression, gradient-boosted decision tree, LASSO-MTL[i], L2-MTL[j], cASO-MTL[k]	CNN features or word vector representations with hand-crafted features	12 stressor events (eg, marriage, financial, illness, school), 6 stressor subjects	30 volunteers manually annotated the stressor events and subjects	10-fold cross validation (F1>0.80)
Weibo	Cheng et al [34]	Suicide risk, depression, anxiety, stress	974 participants' Weibo posts, suicide probability, Weibo suicide communication (WSC), depression, anxiety, and stress.	SVM	Simplified Chinese-LIWC	Suicide risk, emotional distress (depression, anxiety, stress), WSC	Survey and psychological test tools (ie, Suicide Probability Scale, Depression Anxiety Stress Scales-21)	leave-one-out cross-validation: suicide probability (AUC=0.61, P=.04), severe anxiety (AUC=0.75, P<.001)

Data Source	Reference	Mental Health	Data Description	Machine learning				
				Model	Feature	Output	Annotation	Results
Reddit	Gkotsis et al [35]	Bipolar, schizophrenia, anxiety, depression, self-harm, suicide watch, addiction, crippling alcoholism, opiates, autism	1,014,660 posts	CNN, FF[l], linear regression, SVM	Word vector representation (16 vector size)	Mental health	N/A[m]	Accuracy: 91.8% (binary classification task), 79.8% (multiclass classification task)
Facebook, Twitter, Instagram, Reddit	Coppersmith et al [36]	Suicide risk	197,615 posts from 418 users	LSTM with attention	One-hot-vector mapped to pretrained GloVe	Suicide risk	Examining public self-stated data and using data donated through OurDataHelps.org	10-fold cross validation (AUC=0.94)
Online Community - Live Journal	Saha et al [37]	Mental health	620,060 posts from 78,647 users	MTL	Linguistic features of LIWC; topics by LDA[n]	Mental health subreddit (eg, Abuse, Anorexia, Anxiety, Bipolar disorder, Cutting, Death, Drugs, Eating disorders, Insomnia, Pain, Self-injury, and Suicide)	N/A	AUC=0.94 with the community on eating disorders

[a]AUC: area under the curve.

[b]CNN: convolutional neural network.

[c]SVM: support vector machine.

[d]Bi-LSTM: bidirectional long short-term memory.

[e]LASSO: least absolute shrinkage and selection operator.

[f]LIWC: Linguistic Inquiry and Word Count.

[g]DTSVM: decision tree support vector machine.

[h]SNMDD: social network mental disorder detection.

[i]MTL: multitask learning.

[j]l2-MTL: multitask learning considering l2 loss.

[k]cASO-MTL: clustered alternating structure optimization multitask learning.

[l]FF: feed-forward.

[m]N/A: not applicable.

[n]LDA: latent Dirichlet allocation.

In the case of Twitter, Budenz et al [28] analyzed 1,270,902 tweets by searching bipolar and mental health terms. Before using a logistic regression analysis to classify whether a specific tweet was asking for any help or included terms associated with mental illness stigma, they repeatedly performed a series of sentiment analyses on 2047 randomly sampled tweets. The results obtained through 10-fold cross-validation procedures showed an average area under the curve (AUC) of 0.83 when the term-frequency inverses document frequency weighted vector was employed as the input source.

Twitter tweets were also used to predict an association with psychological stressors, as one of the major causes of suicide. To prevent suicidal behaviors [29], the authors retrieved 1,962,766 tweets based on 21 suicide-related keywords, manually annotated labels of a subset of 3263 tweets, and labeled the other 3000 tweets based on ML classifiers. A Twitter corpus–pretrained GloVe vector was employed to convert each token into a vector as an input of the convolutional neural network (CNN). The CNN model achieved an F1-score of 83%, which outperformed SVM, extra trees, and other ML algorithms.

In addition, stress level, as one of the pervasive causes of mental health conditions, was predicted on social media platforms, including Twitter and Facebook [30]. Linguistic characteristics were extracted from a total of 601 users' social media posts using the Linguistic Inquiry and Word Count (LIWC) tool, latent Dirichlet allocation (LDA), and stress lexicon. To predict stress, the authors applied linear regression with regularization methods, and validated the model performance based on sociodemographic variables (eg age, gender, race, income, and education) and social media language using the Pearson correlation coefficient (*r*). The content analysis indicated significant differences in language expressions among social media platforms.

Shuai et al [31] collected posts from both Facebook and Instagram to detect social network mental disorders (SNMD), which include several side effects such as cyber-relationship addiction or net compulsion. They developed and employed SNMD questionnaire items to classify each participant into specific types of SNMD. Among more than 3100 participants, 389 respondents were regarded as having SNMD. Social interaction, personal profile, and duration extracted by each participant's social activity logs were employed as input features of ML models. The proposed model, which was organized by new tensor techniques and latent features, achieved more than 83% accuracy in identifying whether a specific user has SNMD.

One of the notable approaches in this area is a visual-oriented approach. Reece and Danforth [32] employed 43,950 images from 166 Instagram users to detect posts related to depression. Based on the results of the Center for Epidemiologic Studies Depression Scale questionnaire (CES-D), a total of 71 users revealed that they experienced depression. Moreover, both Instagram usernames and history were collected from crowd workers who responded to the CES-D. They extracted metadata (eg, the number of comments, "likes"), color properties (eg, hue, saturation, value), and the total number of faces from the collected photographs to investigate whether users suffer from depression.

Based on the guidelines of De Choudhury et al [38], Reece and Danforth decided to integrate the users' recent posts presented on a specific (single) day rather than using their entire posts. Through a random forest classifier, they achieved a relatively high recall score of identifying the target class at 70% in 100 observations. The results indicated that the photos posted by depressed users were more likely to be bluer, grayer, and darker, and to receive fewer likes. However, as a limitation of the study, they pointed out that depression is a form of general clinical status, indicating a need for fine-tuning the questionnaires for specific diagnosis.

Lin et al [33] collected approximately 1 billion tweets from the Chinese social media platform Weibo, and proposed ML multitask models to detect both stressor events and six subjects. The event was categorized into 12 different labels, including marriage, financial, illness, and school. Each tweet was first labeled as stress-related. The tweets were categorized into one of the stressor events and subject categories by 30 volunteers. The performance of classifying a stressor event or subject was represented with various classifiers such as SVM, softmax regression, and gradient decision. The model performance was not clearly presented; however, it was stated that the F1-score reached over 80% in the event detection task.

A mental illness is often accompanied by another mental illness as a so-called "comorbidity," which refers to the simultaneous presence of one or more mental or physical disorders. From this viewpoint, Saha et al [37] developed a joint learning model, which was generated by multitask learning to simultaneously identify co-occurring social media communities related to mental health with consideration of the correlation between the communities. Based on 620,060 posts of 78,647 users in 247 online communities, 12 major mental health–related topics were employed in the categorization standards from "Live Journal" (eg, Abuse, Anorexia, Anxiety, Bipolar disorder, etc). Using these data, two features were extracted as inputs: language style (from the LIWC) and topics (from the results of LDA). In general, the proposed model outperformed single-task learning [39] and multitask learning [40] for 9 out of 12 and 8 out of 12 categories, respectively. Moreover, the model achieved an AUC of 0.94 with the community on eating disorders.

Cheng et al [34] utilized Weibo data to assess the levels of suicide risk and emotional distress such as depression, anxiety, and stress. For this purpose, the researchers completed an internet survey and gathered 974 respondents' Weibo posts, the scores of mental health (ie, suicide probability, depression, anxiety, and stress) through psychological investigation tools, and Weibo Suicide Communication (WSC), which examined whether respondents had told others that they wanted to commit suicide through Weibo over the past 12 months. SVM was applied for a binary classification of five suicide risk factors (suicide probability, depression, anxiety, stress, and WSC), including 72 linguistic features of Simplified Chinese-LIWC from the respondents' Weibo posts as independent variables. The model efficiently classified the respondents having a high suicide probability (AUC=0.61, P=.04) and severe anxiety (AUC=0.75, P<.001) among those who had WSC with leave-one-out cross-validation procedures.

Gkotsis et al [35] collected user-oriented data from the Reddit community to develop deep-learning models for classifying posts according to mental disorder topics. After an expert panel made a decision on whether a specific post contains mental health–related issues, they collected 10,146,60 posts and extracted 11 mental disorder themes, including a nonmental health conditions. To classify whether a specific post belongs to one of the mental health topics, they employed a CNN model with two parallel classification approaches: binary and multiclass classifications. The word vectors of each token extracted from post texts were used as input. The results of the model demonstrated that the CNN classifier showed 91.8% and 79.8% accuracy in binary and multiclass classification tasks, respectively.

Early estimation of a person's suicide risk is also an important issue in our society. In accordance with this point, Coppersmith et al [36] employed social media data to predict the level of suicide risk using a long short-term memory (LSTM) model. Two different datasets were collected: one from donated data through OurDataHelps.org, which included social media data

of suicide victims (eg, Facebook, Instagram, Twitter), and the other from Harman Dredze [41], which provided Twitter data from users who mentioned their past suicide attempts in tweets. A total of 197,615 posts from 418 users were obtained. Based on the pretrained GloVe embeddings to feed sequences of word vectors into the bidirectional LSTM model, an AUC of 0.94 was achieved through 10-fold cross-validation procedures.

Discussion

This study involved a bibliometric analysis on the publications related to ML and mental health in social media from 2015 to 2020 with two citation databases, WoS and Scopus, as well as a trend review analysis. Although several prior studies have investigated mental illness, the majority of these studies employed both clinical and physical health care approaches. Along with these studies, social media is considered as one of the most important spaces for effectively and efficiently addressing individuals' mental health issues [42]. Furthermore, with rapidly improved ML and big data techniques, both the significance and importance of employing social media and online communications are being consistently emphasized.

Rapidly and consistently increasing publication and citation numbers indicate that there is growing attention and interest in this research area. Among several publication venues, *Lecture Notes in Computer Science* (Scopus) and *Journal of Medical Internet Research* (WoS) were the most productive publication sources in this field. Moreover, Harvard University and Georgia Institute of Technology in the United States, and Tsinghua University and the Chinese Academy of Sciences in China were listed as the most vigorous institutes. For individual researchers, Professor Munmun De Choudhury from Georgia Institute of Technology emerged as the most productive and well-known researcher in this field, with 15 publications to her credit. Regarding publications, using social media data in predicting depression was Prof De Choudhury's first step in this area [38], which allowed her affiliated nation (United States) to be the most productive country in this field. Computer science, medicine, and medical informatics were identified as the core research subjects, along with several other related subjects such as psychology, social science, and neuroscience. This suggests that this research area tends to require integrated or multidisciplinary approaches for gaining a better understanding of each research topic. In addition, the keyword co-occurrence network graph highlighted the representative ML techniques and social media for this multidisciplinary area.

Subsequently, we conducted a trend analysis review on highly cited articles, and notable research trends were identified. The highly cited articles tended to employ user-generated content in diverse forms, including text, images, and other metadata, for specific mental disorders. Because no ground truth labels exist for users who have mental disorders, the majority of studies adopted a crowdsourcing survey with a medical-oriented approach and consideration of the participants' agreements in using their social media accounts [30-32,34]. Moreover, several scholars have employed user-oriented features, including users' demographic profiles and activity logs, in social media (eg, comments, likes) to arrive at both academic and practical contributions [30,31].

Based on the employed approaches with several highly cited articles, three main implications for discussion can be derived. First, the majority of the articles stated privacy and ethical issues as key considerations in using ML for specific mental illness in social media [23,30,32,36]. Although they met both research ethical guidelines and participants' agreements in using their social media data, there were notable adverse reactions from several participants in sharing their social media information [32]. Moreover, compared to other issues in this area, both privacy and ethical issues are considered to be real issues requiring more academic and practical work [23]. Thus, researchers must make efforts to effectively consider these threats, which can negatively affect data providers and leave room for abusing ML techniques. Second, because there is potential for misclassified ground truth data, there should be more detailed and systematic examinations in building the early stage of datasets [28,29,34,36]. Third, because users' expressions on social media can consistently change over time, time-oriented approaches with differing perspectives toward users' activities and expressions should be considered in providing a better understanding in this area [32].

Considering these discussion points, a few limitations of this study remain. Although we employed WoS and Scopus as our subjects, which are both widely used academic databases globally, there can be other medical-oriented databases that may provide more significant academic and practical information. Moreover, considering more recent and applicable statistical or natural language processing techniques (such as exploratory factor analysis [23] or topic modeling [20,23]), future research should aim at obtaining deeper and comprehensive knowledge with more creative and significant approaches through various data sources of each article.

With the increase of AI applications in real-word health care settings [1], there have been numerous attempts to overcome the limitations of offline consultations such as wearable fitness trackers [43], mobile health apps [44], and conversational agents for patients' mental health and wellness [1,45]. Moreover, since the use of social media has been widely adopted in health care [46], we believe that our analysis may trigger all stakeholders to further consider how to employ ML approaches toward mental health in social media. In addition, when applying social media data to clinical settings, there is a need to address different characteristics of social media platforms by utilizing the substantial research background.

Acknowledgments

This work was supported by an Institute of Information and Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (Ministry of Science and ICT; No. 2020-0-00990, Platform Development and Proof of High

Trust & Low Latency Processing for Heterogeneous Atypical Large Scaled Data in 5G-IoT Environment). This research was also supported by the National Research Foundation of Korea funded by the Korean Government (NRF-2020R1C1C1004324).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete publication lists.
[XLSX File (Microsoft Excel File), 64 KB - jmir_v23i3e24870_app1.xlsx]

References

1. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
2. Kahan J. Mobilizing AI for health to fight against COVID-19. Microsoft blogs. 2020 Apr 09. URL: https://blogs.microsoft.com/on-the-issues/2020/04/09/ai-for-health-covid-19 [accessed 2020-09-18]
3. Castelo M. The Future of Artificial Intelligence in Healthcare. HealthTech. 2020 Feb 26. URL: https://healthtechmagazine.net/article/2020/02/future-artificial-intelligence-healthcare [accessed 2020-09-18]
4. Hao K. Doctors are using AI to triage covid-19 patients. The tools may be here to stay. MIT Technology Review. 2020 Apr 23. URL: https://www.technologyreview.com/2020/04/23/1000410/ai-triage-covid-19-patients-health-care/ [accessed 2020-09-18]
5. Yu K, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018 Oct;2(10):719-731. [doi: 10.1038/s41551-018-0305-z] [Medline: 31015651]
6. Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: A scoping review. PLoS One 2019;14(2):e0212356 [FREE Full text] [doi: 10.1371/journal.pone.0212356] [Medline: 30779785]
7. World Mental Health Day: an opportunity to kick-start a massive scale-up in investment in mental health. World Health Organization. 2020 Aug 27. URL: https://www.who.int/news-room/detail/27-08-2020-world-mental-health-day-an-opportunity-to-kick-start-a-massive-scale-up-in-investment-in-mental-health [accessed 2020-09-18]
8. GBD 2017 Disease Injury Incidence Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
9. Batterham PJ, Calear AL. Preferences for internet-based mental health interventions in an adult online sample: findings from an online community survey. JMIR Ment Health 2017 Jun 30;4(2):e26 [FREE Full text] [doi: 10.2196/mental.7722] [Medline: 28666976]
10. Brunette M, Achtyes E, Pratt S, Stilwell K, Opperman M, Guarino S, et al. Use of smartphones, computers and social media among people with SMI: opportunity for intervention. Community Ment Health J 2019 Aug 8;55(6):973-978. [doi: 10.1007/s10597-019-00431-7] [Medline: 31175518]
11. Naslund JA, Bondre A, Torous J, Aschbrenner KA. Social media and mental health: benefits, risks, and opportunities for research and practice. J Technol Behav Sci 2020 Sep;5(3):245-257. [doi: 10.1007/s41347-020-00134-x] [Medline: 33415185]
12. Breland JY, Quintiliani LM, Schneider KL, May CN, Pagoto S. Social media as a tool to increase the impact of public health research. Am J Public Health 2017 Dec;107(12):1890-1891. [doi: 10.2105/AJPH.2017.304098] [Medline: 29116846]
13. Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. Chichester, UK: John Wiley & Sons Ltd and The Cochrane Collaboration; 2008:187-241.
14. Seabrook EM, Kern ML, Rickard NS. Social networking sites, depression, and anxiety: a systematic review. JMIR Ment Health 2016 Nov 23;3(4):e50 [FREE Full text] [doi: 10.2196/mental.5842] [Medline: 27881357]
15. Pham MT, Rajić A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res Synth Methods 2014 Dec;5(4):371-385 [FREE Full text] [doi: 10.1002/jrsm.1123] [Medline: 26052958]
16. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med 2019 Jul;49(9):1426-1448. [doi: 10.1017/S0033291719000151] [Medline: 30744717]
17. Chancellor S, Baumer EPS, De Choudhury M. Who is the human in human-centered machine learning? The case of predicting mental health from social media. 2019 Nov 07 Presented at: ACM on Human-Computer Interaction, 3 (CSCW); November 2019; Texas, USA p. 1-32. [doi: 10.1145/3359249]
18. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digit Med 2020;3:43. [doi: 10.1038/s41746-020-0233-7] [Medline: 32219184]
19. Pritchard A. Statistical bibliography or bibliometrics. J Document 1969;25(4):348-349.

20. Chen X, Wang S, Tang Y, Hao T. A bibliometric analysis of event detection in social media. Online Inf Rev 2019 Feb 11;43(1):29-52. [doi: 10.1108/oir-03-2018-0068]
21. Zyoud SH, Sweileh WM, Awang R, Al-Jabi SW. Global trends in research related to social media in psychology: mapping and bibliometric analysis. Int J Ment Health Syst 2018;12:4 [FREE Full text] [doi: 10.1186/s13033-018-0182-6] [Medline: 29387147]
22. dos Santos BS, Steiner MTA, Fenerich AT, Lima RHP. Data mining and machine learning techniques applied to public health problems: A bibliometric analysis from 2009 to 2018. Comput Indust Eng 2019 Dec;138:106120. [doi: 10.1016/j.cie.2019.106120]
23. Tran BX, McIntyre RS, Latkin CA, Phan HT, Vu GT, Nguyen HLT, et al. The current research landscape on the artificial intelligence application in the management of depressive disorders: a bibliometric analysis. Int J Environ Res Public Health 2019 Jun 18;16(12):2150 [FREE Full text] [doi: 10.3390/ijerph16122150] [Medline: 31216619]
24. Wang J, Deng H, Liu B, Hu A, Liang J, Fan L, et al. Systematic evaluation of research progress on natural language processing in medicine over the past 20 years: bibliometric study on PubMed. J Med Internet Res 2020 Jan 23;22(1):e16816 [FREE Full text] [doi: 10.2196/16816] [Medline: 32012074]
25. Kim J, Lee J, Park E, Han J. A deep learning model for detecting mental illness from user content on social media. Sci Rep 2020 Jul 16;10(1):11846. [doi: 10.1038/s41598-020-68764-y] [Medline: 32678250]
26. Lee P, Su H. Investigating the structure of regional innovation system research through keyword co-occurrence and social network analysis. Innovation 2014 Dec 17;12(1):26-40. [doi: 10.5172/impp.12.1.26]
27. Bardus M, El Rassi R, Chahrour M, Akl EW, Raslan AS, Meho LI, et al. The use of social media to increase the impact of health research: systematic review. J Med Internet Res 2020 Jul 06;22(7):e15607 [FREE Full text] [doi: 10.2196/15607] [Medline: 32628113]
28. Budenz A, Klassen A, Purtle J, Yom Tov E, Yudell M, Massey P. Mental illness and bipolar disorder on Twitter: implications for stigma and social support. J Ment Health 2020 Apr;29(2):191-199. [doi: 10.1080/09638237.2019.1677878] [Medline: 31694433]
29. Du J, Zhang Y, Luo J, Jia Y, Wei Q, Tao C, et al. Extracting psychiatric stressors for suicide from social media using deep learning. BMC Med Inform Decis Mak 2018 Jul 23;18(Suppl 2):43 [FREE Full text] [doi: 10.1186/s12911-018-0632-8] [Medline: 30066665]
30. Guntuku S, Buffone A, Jaidka K, Eichstaedt J, Ungar L. Understanding and measuring psychological stress using social media. 2019 Jul Presented at: Thirteenth International AAAI Conference on Web and Social Media; July 2019; Munich, Germany p. 214-225.
31. Shuai H, Shen C, Yang D, Lan Y, Lee W, Yu P, et al. A comprehensive study on social network mental disorders detection via online social media mining. IEEE Trans Knowl Data Eng 2018 Jul 1;30(7):1212-1225. [doi: 10.1109/TKDE.2017.2786695]
32. Reece AG, Danforth CM. Erratum to: Instagram photos reveal predictive markers of depression. EPJ Data Sci 2017 Sep 5;6(1):1-12. [doi: 10.1140/epjds/s13688-017-0118-4]
33. Lin H, Jia J, Nie L, Shen G, Chua T. What does social media say about your stress? 2016 Jul Presented at: Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16); July 2016; New York, USA p. 3775-3781 URL: https://dl.acm.org/doi/abs/10.5555/3061053.3061147
34. Cheng Q, Li TM, Kwok C, Zhu T, Yip PS. Assessing suicide risk and emotional distress in Chinese social media: a text mining and machine learning study. J Med Internet Res 2017 Jul 10;19(7):e243 [FREE Full text] [doi: 10.2196/jmir.7276] [Medline: 28694239]
35. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJP, Dobson RJB, et al. Characterisation of mental health conditions in social media using Informed Deep Learning. Sci Rep 2017 Mar 22;7:45141. [doi: 10.1038/srep45141] [Medline: 28327593]
36. Coppersmith G, Leary R, Crutchley P, Fine A. Natural language processing of social media as screening for suicide risk. Biomed Inform Insights 2018;10:1178222618792860 [FREE Full text] [doi: 10.1177/1178222618792860] [Medline: 30158822]
37. Saha B, Nguyen T, Phung D, Venkatesh S. A Framework for Classifying Online Mental Health-Related Communities With an Interest in Depression. IEEE J Biomed Health Inform 2016 Jul;20(4):1008-1015. [doi: 10.1109/JBHI.2016.2543741] [Medline: 27008680]
38. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. 2013 Apr Presented at: SIGCHI conference on human factors in computing systems; April 2013; Paris, France p. 3267-3276. [doi: 10.1145/2470654.2466447]
39. Liu J, Chen J, Ye J. Large-scale sparse logistic regression. 2009 Jun Presented at: 15th ACM SIGKDD international conference on knowledge discovery and data mining; June 2009; Paris, France p. 547-556 URL: https://dl.acm.org/doi/abs/10.1145/1557019.1557082
40. Liu J, Ji S, Ye J. Multi-task feature learning via efficient ‘2, 1-norm minimization. 2009 Jun Presented at: Twenty-Fifth Conference on Uncertainty in Artificial Intelligence; June 2009; Quebec, Canada p. 339-348 URL: https://dl.acm.org/doi/10.5555/1795114.1795154

41. Harman G, Dredze M. Measuring post traumatic stress disorder in Twitter. 2014 May Presented at: Eighth International AAAI Conference on Weblogs and Social Media; May 2014; Michigan, USA.
42. De Choudhury M. Role of social media in tackling challenges in mental health. 2013 Oct Presented at: 2nd international workshop on socially-aware multimedia; October 2013; Barcelona, Spain p. 49-52. [doi: 10.1145/2509916.2509921]
43. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res 2018 Mar 22;20(3):e110 [FREE Full text] [doi: 10.2196/jmir.9157] [Medline: 29567635]
44. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
45. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
46. Ukoha C, Stranieri A. Criteria to measure social media value in health care settings: narrative literature review. J Med Internet Res 2019 Dec 16;21(12):e14684 [FREE Full text] [doi: 10.2196/14684] [Medline: 31841114]

Abbreviations

AI: artificial intelligence
AUC: area under the curve
CES-D: Center for Epidemiologic Studies Depression Scale
CNN: convolutional neural network
HCML: human-centered machine learning
LDA: latent Dirichlet allocation
LIWC: Linguistic Inquiry and Word Count
LSTM: long short-term memory
ML: machine learning
SNMD: social network mental disorders
SVM: support vector machine
WoS: Web of Science
WSC: Weibo Suicide Communication

Edited by R Kukafka; submitted 08.10.20; peer-reviewed by H Chang, A Raman; comments to author 24.10.20; accepted 17.02.21; published 08.03.21.

Please cite as:
Kim J, Lee D, Park E
Machine Learning for Mental Health in Social Media: Bibliometric Study
J Med Internet Res 2021;23(3):e24870
URL: https://www.jmir.org/2021/3/e24870
doi:10.2196/24870
PMID:33683209

Original Paper

Parents' Perspectives on Using Artificial Intelligence to Reduce Technology Interference During Early Childhood: Cross-sectional Online Survey

Jill Glassman[1], MSW, PhD; Kathryn Humphreys[2], PhD; Serena Yeung[3], PhD; Michelle Smith[4], BA; Adam Jauregui[1], MSc; Arnold Milstein[1], MD; Lee Sanders[4], MD

[1]Clinical Excellence Research Center, School of Medicine, Stanford University, Stanford, CA, United States

[2]Department of Psychology and Human Development, Vanderbilt University, Nashville, TN, United States

[3]Department of Biomedical Data Science, School of Medicine, Stanford University, Stanford, CA, United States

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

Corresponding Author:
Jill Glassman, MSW, PhD
Clinical Excellence Research Center
School of Medicine
Stanford University
365 Lasuen Street
#308
Stanford, CA, 94305
United States
Phone: 1 8314195302
Email: jill.r.glassman@stanford.edu

Abstract

Background: Parents' use of mobile technologies may interfere with important parent-child interactions that are critical to healthy child development. This phenomenon is known as technoference. However, little is known about the population-wide awareness of this problem and the acceptability of artificial intelligence (AI)–based tools that help with mitigating technoference.

Objective: This study aims to assess parents' awareness of technoference and its harms, the acceptability of AI tools for mitigating technoference, and how each of these constructs vary across sociodemographic factors.

Methods: We administered a web-based survey to a nationally representative sample of parents of children aged ≤5 years. Parents' perceptions that their own technology use had risen to potentially problematic levels in general, their perceptions of their own parenting technoference, and the degree to which they found AI tools for mitigating technoference acceptable were assessed by using adaptations of previously validated scales. Multiple regression and mediation analyses were used to assess the relationships between these scales and each of the 6 sociodemographic factors (parent age, sex, language, ethnicity, educational attainment, and family income).

Results: Of the 305 respondents, 280 provided data that met the established standards for analysis. Parents reported that a mean of 3.03 devices (SD 2.07) interfered daily in their interactions with their child. Almost two-thirds of the parents agreed with the statements "I am worried about the impact of my mobile electronic device use on my child" and "Using a computer-assisted coach while caring for my child would help me notice more quickly when my device use is interfering with my caregiving" (187/281, 66.5% and 184/282, 65.1%, respectively). Younger age, Hispanic ethnicity, and Spanish language spoken at home were associated with increased technoference awareness. Compared to parents' perceived technoference and sociodemographic factors, parents' perceptions of their own problematic technology use was the factor that was most associated with the acceptance of AI tools.

Conclusions: Parents reported high levels of mobile device use and technoference around their youngest children. Most parents across a wide sociodemographic spectrum, especially younger parents, found the use of AI tools to help mitigate technoference during parent-child daily interaction acceptable and useful.

(J Med Internet Res 2021;23(3):e19461) doi:10.2196/19461

KEYWORDS

parenting; digital technology; mobile phone; child development; artificial intelligence

Introduction

Technology Interference in Responsive Parenting

Parent and caregiver behaviors that foster healthy, nurturing environments during early childhood produce significant, cost-effective benefits across the life course well into adulthood [1-3]. Exposure to such environments during the infant and toddler years—including verbally rich interactions, affective-emotional caregiver behaviors (eg, returning the baby's gaze, positive touch), and the maintenance of the child's focus of interest—is associated with positive cognitive, linguistic, physical, and socioemotional health outcomes [4-6]. However, over the past decade, a new threat to these environments has emerged: the use of mobile technologies by parents during active caregiving [7-9]. In fact, recent studies suggest that this parenting technoference is associated with reduced adult responsiveness and early child behavior problems [8-10]. Furthermore, these risks to subsequent lifelong health may be highest in low-income households, which disproportionately rely on mobile technology for access to the web and social support [11].

Early Childhood Parenting Interventions

Interventions most effective in supporting long-term outcomes for children focus on helping parents foster nurturing environments during early childhood [12-15]. Many of these interventions include technological support for wider dissemination or the use of digital technologies as a mode of implementation.For example, the use of audio-detection technology to improve the quality of early childhood literacy environments is used by the Language ENvironment Analysis (LENA) word count device to monitor the linguistic development of young children and provide feedback to parents [16-19]. In another example, the Video Interaction Project provides live coaching feedback to parents based on 5-minute videos of parent-child interactions during regularly scheduled well-child visits [20,21]. However, the moderating effects of such parenting interventions on technoference have not yet been studied. Furthermore, human, financial, and technological constraints limit their scalability.

Augmenting Parenting Interventions With Artificial Intelligence–Based Tools

Emergent artificial intelligence (AI) capabilities—including noninvasive visual and audio monitoring tools that use trained algorithms to automatically detect certain human behaviors—may enable more efficient and scalable behavioral observation to augment interventions that support key nurturing parenting behaviors. In fact, in both research and commercial applications, this has proven to be successful; wearable devices support adult lifestyle changes (eg, physical activity), and more recently in hospital settings, computer vision is being employed to promote adherence to patient care protocols such as hand hygiene and patient mobilization [22,23]. Given the increasing presence of mobile devices in the parent-child environment, AI-based technology may be an important tool for identifying and implementing interventions that help parents reduce technology interference in their relationships with their young children. However, little is known about the acceptability of using AI tools to further identify and mitigate the effects of parent technoference.

Acceptability of New Technology

One of the most common models used to understand and measure the acceptability of new technology in general, and in health care in particular, is the technology acceptance model (TAM). Originally developed in the context of the adoption of workplace technology tools, the TAM posits that the perceived usefulness (PU) and the ease of use of a new technology determine its acceptability, which in turn influences the intention to use and actual use of the new technology [24-26]. In addition, the literature on the acceptability of health- and behavior-related technology, such as wearables, telemedicine, and mobile health apps, emphasizes that individual sociodemographic factors such as age, gender, and income also play important roles and should be considered in any early technology-based intervention development processes [27-30]. More specifically, there is some evidence that parental engagement in technology-assisted parenting interventions may be higher for younger parents, with mixed findings on whether lower- or higher-income parents are more likely to use these interventions [14].

A review of parental engagement in psychological interventions for their children indicated that beliefs about the severity of the problem may influence engagement with the treatment [31]. Furthermore, the widely used Health Belief Model (HBM) suggests that parents who are more aware of their susceptibility to the problem of technoference would be more accepting of solutions for addressing it [32]. Similar to the TAM, the HBM also suggests that the more useful parents perceive AI-based tools, the more likely they will be to adopt them. However, these theories have not yet been tested in the emerging context of technoference and AI-based tools to reduce it.

Study Goals

Given these gaps, the overall purpose of this exploratory study is to investigate the magnitude and potential determinants of the acceptability of AI-based tools for assisting parents in reducing parenting technoference. In particular, the goals were to assess (1) parental perceptions of their own technoference and its harms, (2) the acceptability of AI-based tools to reduce the negative effects of technoference, (3) the variation in acceptability and technoference across sociodemographic factors, and (4) the relative importance of parents' perceptions of their own technoference and sociodemographic factors in relation to acceptability.

Methods

Design

We conducted a cross-sectional observational study using a population-based survey on the web of US parents of children

aged ≤5 years. The survey was developed for the Amazon TurkPrime (now called CloudResearch) platform for recruitment and administration through the Prime Panels service. Prime Panels is a research participant recruitment platform developed as a step up from Amazon Mechanical Turk (MTurk), in that it includes a set of validated attention and language comprehension screening questions [33,34]. The survey was pilot tested for comprehension on a convenience sample of 4 mothers from the San Francisco Bay Area. The survey was then translated into the Spanish language by a trained, bilingual research associate and back-translated from Spanish to English by another independent, trained bilingual research associate; differences between the original and back-translated versions were resolved in Spanish by a third bilingual research coordinator. Participants could choose to complete the survey in Spanish or English. Recruitment began on May 24, 2019, and ended on June 13, 2019. All procedures were approved by the Stanford Human Subjects Research Office (Institutional Review Board protocol number 50428).

Participants

Eligibility criteria were adults aged >18 years, with primary caregiving responsibility for at least one child aged <5 years in the household. Exclusion criteria included the inability to read English or Spanish or completion of the survey outside the United States. Using the Prime Panels sampling frame, we aimed to recruit a sample representative of the US population, with >25% respondents self-reporting an underrepresented minority status (Hispanic, Black, and Asian or Pacific Islander), >25% with educational attainment less than a college degree, and >15% monolingual Spanish speakers.

Survey Items

The survey consisted of items adapted from existing validated surveys covering the following domains: parenting and child behavior priorities, parents' perceptions of their technology use, parents' perceptions of their technology use in the presence of their child (technoference), the acceptability of AI-based tools to help reduce parenting technoference, and sociodemographic factors. The full survey can be found in Multimedia Appendix 1.

Primary Outcome

The primary outcome was technology acceptance, defined as the acceptability and perceived utility of an AI-based tool for reducing parenting technoference and measured by averaging responses to the following items having 6-point Likert scale response options (strongly disagree to strongly agree): (1) "Using a computer-assisted coach while caring for my child would help me be more aware of my device use around my child," (2) "Using a computer-assisted coach while caring for my child would improve my interactions with my child," (3) "Using a computer-assisted coach while caring for my child would help me be a better parent," (4) "Using a computer-assisted coach while caring for my child would help me notice more quickly when my device use is interfering with my caregiving," (5) "Using a computer-assisted coach while caring for my child would help me keep my attention focused on my child," and (6) "Using a computer-assisted coach while caring for my child would be useful to me." The items were preceded by the following statement:

> *Some electronic devices are currently being designed to HELP you have a better connection with your child—by coaching you or giving you meaningful, real-time feedback. Imagine such a "computer-assisted coach," which you could use in your home to get feedback on your use of electronic devices while caring for your child. The computer-assisted coach would automatically analyze computer vision and other data to provide the feedback. Whenever you want, you could turn this computer-assisted coach on or off.*

These items were adapted from the PU scale of the TAM (TAM-PU) [24-26,35].

Secondary Outcomes and Independent Variables

Problem technology use was defined as the extent to which parents perceived that their mobile technology use had risen to a problematic level and was measured using a previously validated scale [9], which averaged responses to the following three items: "When my mobile electronic device alerts me to indicate new messages, I cannot resist checking them"; "I often think about calls or messages I might receive on my mobile phone"; and "I feel like I use my mobile phone too much." The response options for all three items were on a 6-point scale ranging from strongly disagree to strongly agree. Finally, perception of parenting technoference was measured based on a previously validated index [9], which summed the count of dichotomized responses to the following questions across each of the 6 device types (television, computer, smartphone, tablet, other handheld devices [eg, iPod], and video game device): "In a typical day, how many times does each of the following devices interrupt a conversation or activity between you and your child?" Possible responses in our survey used a 5-point scale (never, 1 time, 2 times, 3 times, and 4 or more times) and were dichotomized to 0 times versus 1 or more times.

Sociodemographic characteristics included self-report of age, sex, race or ethnicity, language spoken at home (English or Spanish), marital status, number of children, education level, working status, and income level. A proxy measure of the geographic region in which the participant resided was derived from the longitude and latitude values of the survey respondent's computer captured by Amazon TurkPrime. The geographic region was categorized into the 4 US census regions defined by the West, South, Northeast, and Midwest.

Data Analysis

Survey responses were first analyzed to assess their distribution. Data quality assessments were conducted to identify speeders (those who answer unreasonably fast) and straightliners (those who answer with identical values for each survey item in a block). We defined speeders as anyone who finished the survey at an average speed of less than 2 seconds per question, and we defined straightliners as anyone who had a standardized scale point variation value of −3.79 [36,37]. The internal consistency of the problem technology use and technology acceptance scale was assessed using the Cronbach α. The Cronbach α was not

calculated for parenting technoference, given that it is a sum of items not necessarily expected to be related [9].

Descriptive statistics were used to address goals 1 and 2, and bivariate analyses were conducted to address goal 3. Given that goal 4 was simply to explore relative, independent associations between independent variables represented by problem technology use, parenting technoference, and sociodemographic factors and the dependent variable technology acceptance rather than to determine multiple causal pathways, we selected multiple regression analysis as the most straightforward, parsimonious way to accomplish this goal [38]. A screening criterion of $P<.15$ was used to determine the sociodemographic characteristics sufficiently associated with the technology acceptance outcome to be entered into the multiple regression models. This *P* value was selected for the purposes of screening as it has been shown that model-building strategies using a .05 criterion often result in the omission of covariates known to be important [39,40].

Multiple linear regression was conducted on full (all sociodemographic covariates) and reduced (only covariates meeting the screening criteria) models. Effect size was estimated using Cohen *d* values or odds ratios (continuous and categorical independent variables) and η^2 values, representing the proportion of total variance in the technology acceptance outcome explained by each independent variable [41].

Mediation analysis was conducted to further investigate the relative importance of the roles played by problem technology use in general and the more specific parenting technoference measure in influencing technology acceptance. The purpose of the mediation analysis was to help assess whether to focus potential behavior change levers on perceived excessive technology use in general and/or technology interference in parenting specifically. All analyses were conducted using the R version 3.5.3 (R Core Team).

Results

Sample

The initial sample of consenting participants meeting the inclusion criteria consisted of 305 survey respondents. Data quality analyses to identify and remove speeders and straightliners resulted in a final analytic sample of 280 observations.

The mean age of the respondents in the final analytic sample was 33 (SD 8) years. In total, 79.2% (222/280) of respondents were female (Table 1), 14.8% (14/270) self-identified as Hispanic, 8.9% (24/270) self-identified as Black, and 5.9% (16/270) self-identified as Asian. Approximately one-third (81/280, 28.9%) of the participants reported less than a high school education, and 13.9% (39/280) spoke a language other than English at home.

Table 1. Descriptive statistics for the study sample (N=280).

Characteristics	Values[a,b]
Age (years), mean (SD)	33 (8)
Gender, n (%)	
Female	222 (79.2)
Male	57 (20.4)
Other	1 (0.4)
Race, n (%)	
White	184 (68.1)
Black	24 (8.9)
Hispanic	40 (14.8)
Asian	16 (5.9)
Other	6 (2.2)
Language spoken at home, n (%)	
English	241 (86.1)
Other	39 (13.9)
Children, n (%)	
1	108 (38.6)
>1	172 (61.4)
Education, n (%)	
Less than high school	81 (28.9)
Some college	65 (23.2)
Higher than or equal to a college degree	134 (47.9)
Income (US $), n (%)	
<25,000	57 (20.4)
25,000-49,999	68 (24.3)
50,000-74,999	65 (23.2)
75,000-100,000	49 (17.5)
>100,000	41 (14.6)
Geographic area, n (%)	
Midwest	57 (20.4)
Northeast	49 (17.5)
South	117 (41.8)
West	57 (20.4)

[a]Total sample size differs across characteristics due to missing values.

[b]Some percentages add up to slightly less than 100% because of missing values.

Descriptive Statistics and Bivariate Associations

Technology Acceptance

Approximately two-thirds of the respondents (184/282, 65.2% and 175/283, 61.8%, respectively) agreed with the statements "Using a computer-assisted coach while caring for my child would help me notice more quickly when my device use is interfering with my caregiving" and "Using a computer-assisted coach while caring for my child would help me be more aware of my device use around my child." The internal consistency of the technology acceptance scales was high, with a Cronbach α of .94 (Table 2). The mean level for the technology acceptance outcome was 3.53 (SD 1.29) on a 6-point scale, where higher values represent higher levels of acceptance.

Table 2. Bivariate associations between sociodemographic characteristics and technology acceptance, problem technology use, and parenting technoference.

Measure	Problem technology use	*P* value	Parenting technoference	*P* value	Technology acceptance	*P* value
Cronbach α	.80	N/A[a]	N/A	N/A	.94	N/A
Overall, mean (SD)	3.72 (1.32)	N/A	3.03 (2.07)	N/A	3.53 (1.29)	N/A
Age of parent (years; ρ)	−0.12	.04[b,c]	−0.12	.05	−0.11	.07
Sex, mean (SD)		.20		.99		.04[c]
Female	3.67 (1.28)		3.03 (2.03)		3.46 (1.25)	
Male	3.89 (1.51)		3.05 (2.25)		3.79 (1.42)	
Education, mean (SD)		.17		.13		.11
Less than high school	3.51 (1.39)		3.23 (2.12)		3.34 (1.45)	
Some college	3.68 (1.22)		2.58 (2.07)		3.42 (1.12)	
Higher than or equal to a college degree	3.86 (1.33)		3.11 (2.04)		3.69 (1.25)	
Race or ethnicity, mean (SD)		.03[c]		.02[c]		.23
White non-Hispanic	3.63 (1.36)		2.82 (2.0)		3.43 (1.34)	
White Hispanic	3.67 (1.32)		4.0 (2.11)		3.78 (1.25)	
Black	4.14 (1.02)		2.96 (2.16)		3.72 (1.13)	
Other[d]	4.32 (1.21)		3.23 (2.18)		3.89 (1.16)	
Language at home, mean (SD)		.50		.005[c]		.92
English	3.69 (1.33)		2.88 (2.03)		3.52 (1.3)	
Other	3.88 (1.27)		3.92 (2.16)		3.6 (1.2)	
Income (US $), mean (SD)		.11		.91		.14
<25,000	3.93 (1.08)		2.91 (1.98)		3.54 (1.29)	
25,000-<49,999	3.45 (1.38)		2.96 (2.28)		3.24 (1.15)	
50,000-<74,999	3.64 (1.42)		3.2 (2.01)		3.5 (1.39)	
75,000-<100,000	3.64 (1.12)		2.98 (2.11)		3.79 (1.15)	
≥100,000	4.06 (1.54)		3.07 (1.97)		3.7 (1.44)	
Children at home, mean (SD)		.75		.90		.23
1	3.66 (1.41)		3.06 (2.1)		3.64 (1.24)	
>1	3.75 (1.27)		3.01 (2.06)		3.46 (1.31)	
Geographic region, mean (SD)		.78		.76		.96
Midwest	3.76 (1.19)		2.91 (1.96)		3.53 (1.29)	
Northeast	3.86 (1.23)		3.02 (2.16)		3.54 (1.28)	
South	3.66 (1.37)		2.95 (2.08)		3.48 (1.31)	
West	3.67 (1.45)		3.3 (2.13)		3.6 (1.27)	

[a]N/A: not applicable.

[b]*P* values calculated using the t-test, the Mann-Whitney test (for 2 categories) or the Kruskal-Wallis test (for >2 categories).

[c]*P* values of <.05.

[d]Includes Asian, Pacific Islander, and other. Others excluded from test because of sparsity.

Problem Technology Use

In total, 62.5% (177/283) of the parents agreed with the statement “When my mobile electronic device alerts me to indicate new messages, I cannot resist checking them.” The problem technology use scale had high internal consistency, with a Cronbach α of .8 (Table 2). The mean level of perceived problematic mobile device use in general for parents in our

sample was 3.72 (SD 2.07) on a 6-point scale where a higher score was more problematic (Table 2).

Parenting Technoference

Around 75.6% (214/283) of the parents reported that smartphones interfered in their parent-child interactions at least once daily. Parents in our sample reported a mean of 3.03 devices (SD 2.07) interfering in their interactions with their child on a daily basis (Table 2).

Sociodemographic Factors

Age, sex, race/ethnicity, and language in the home were all significantly associated with at least one of the primary measures (Table 2). In particular, younger parents perceived that they had a greater level of problem technology use in general and parenting technoference in particular and were more accepting of a technology-based tool to help reduce these problems. Males were slightly more accepting of technology-based solutions to reduce parenting technoference. Parents identifying as Black reported higher levels of problem technology use in general, whereas those identifying as Hispanic reported higher levels of parenting technoference. Parents who reported that they speak a language other than English (overwhelmingly Spanish) at home had higher levels of perceived parenting technoference. There was no significant association between parents' reported technology use and their educational attainment, annual income, or geographic region (Table 2).

Multiple Regression and Mediation Analyses

Controlling for sociodemographic factors, the association between parents' acceptance of technology-based tools to combat technoference in parent-child interactions and their perceptions of their own problem technology use was high, with a large Cohen *d* effect size [41] of 0.98 ($P<.001$) and an η^2 value of 27% of the total variance explained (Table 3). The effect size for the association between technology acceptance and parenting technoference was 0.51 ($P<.001$), with an η^2 value of 6% of the variance explained (Table 3). None of the sociodemographic factors measured were significantly associated with technology acceptance in the regression models, once parents' own perceptions of their problem technology use and parenting technoference were accounted for. The results of mediation analyses (Figure 1) indicate that the relationship between problem technology use and technology acceptance did not have a strong, indirect component that was mediated through the more specific parent technoference construct. The magnitude of the indirect effect was only 0.08, compared with 0.42 for the direct effect of problem technology use on technology acceptance, with 16% of the total effect mediated through technoference.

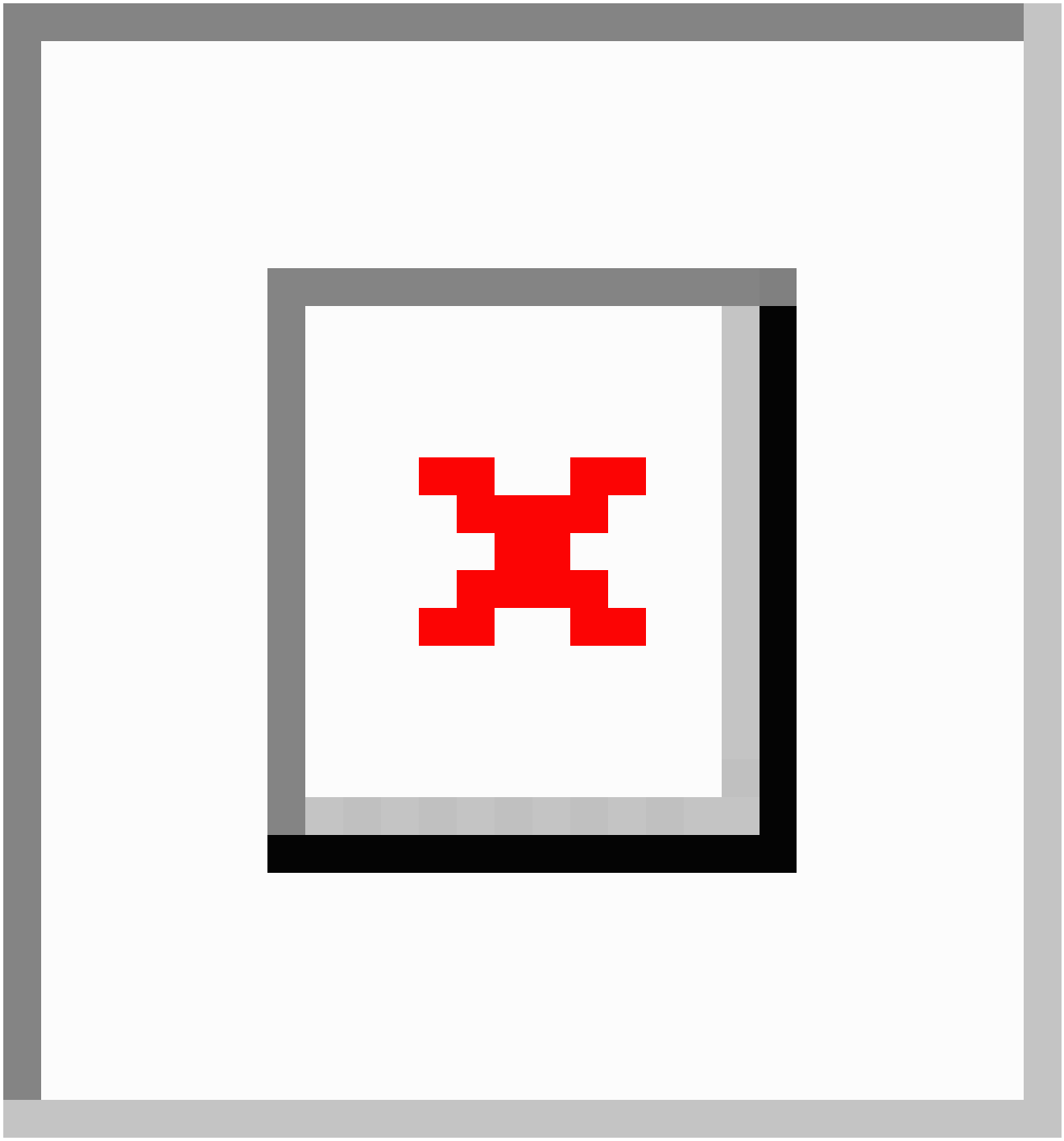
Table 3. Results of the final reduced regression model with technology acceptance as the outcome.

Variable	Regression estimate	Cohen *d*	OR (95% CI)	*P* value	Partial η^2 value (%)
Problem technology use	0.42	0.98	N/A[a]	<.001[b]	27
Parenting technoference	0.14	0.51	N/A	<.001[b]	6
Age of parent (years)	−0.006	−0.09	N/A	.48	2
Sex				.18	2
Female	N/A	N/A	N/A		
Male	0.22		1.24 (0.9, 1.71)		
Education				.30	3
Less than college degree	N/A	N/A	N/A		
College degree	0.15	N/A	1.16 (0.87, 1.54)		
Income (US $)				.33	2
<25,000	N/A	N/A	N/A		
25,000-<50,000	−0.11	N/A	0.9 (0.61, 1.32)		
50,000-<75,000	−0.03	N/A	0.97 (0.65, 1.45)		
75,000-<100,000	0.3	N/A	1.35 (0.88, 2.07)		
>100,000	−0.03	N/A	0.97 (0.60, 1.56)		

[a]N/A: not applicable.

[a]$P<.05$.

Figure 1. Results of the mediation analysis. Mean effect sizes and SDs are shown.



Discussion

In this national survey of parents of young children, we found high levels of acceptance for AI-based coaching tools designed to help reduce threats from parenting technoference to the nurturing environments needed in early childhood to ensure a healthy developmental trajectory to adulthood. This acceptance was most strongly related in regression models to 2 factors: perceived problematic parent technology use in general and perceived level of technoference in the parent-child relationship in particular. In comparison, none of the sociodemographic factors we examined explained a statistically significant portion of the variance in the acceptance of AI-based coaching tools once the factors measuring perceptions of their own susceptibility to the problem of technoference were considered.

Although we found no studies to date on the acceptability of AI-based parenting tools to compare with our results, our findings are somewhat consistent with those of related literature on the acceptability of AI tools in general and for technology-based tools in health care in particular. For example, in one recent study of factors influencing the use of in-home voice assistants, only household size, but not age and gender, influenced usage. In a study of physical activity–monitoring wearable devices, females were found to be slightly more accepting (higher PU) than males [42]. A review of technology-based parenting interventions concluded that although younger parents may be more open to parenting

interventions, they may appeal to a wide range of income levels [14].

Although we were unable to find published studies reporting on the direct relationship between parenting technoference and age, gender, race, and income, the finding that younger parents had higher perceived levels of problem technology use and parenting technoference is consistent with the findings from studies on the relationship between smartphone addiction and age [43,44]. Given that technoference and problem technology use are higher in younger parents, whom we also found to be more open to AI-based tools to mitigate technoference, we suggest that young parents should be the focus of future research and development in this area. However, the lack of a significant association between perceived device use and acceptability and education or income level suggests that the ubiquity of mobile devices is making parenting technoference a broad public health problem requiring scalable solutions.

The finding that parents' perceptions of their own technoference and problematic technology use explained by far the largest proportion of variance in acceptability of AI-based tools to reduce parenting technoference may reflect the robustness of the HBM, in which our technoference measure corresponds to parents' perceptions of their own susceptibility to this problem [32]. Thus, helping parents increase their awareness of their own technoference should be considered as a potential lever for increasing the acceptability of scalable AI-based interventions to mitigate this threat to early childhood development.

Finally, the finding from regression and mediation analyses that the most important factor influencing acceptability of AI tools to reduce technoference was parents' perceptions of general technology overuse compared with their perceptions of how much their technology interfered with their parenting warrants further investigation. AI-based coaching tools or any technology-augmented tools to help reduce parent technoference might be able to focus simply on reducing parents' overall problematic technology use. In addition, further research is needed to uncover other characteristics of parents who view themselves as having problematic digital device use, given that we did not identify strong relationships between the sociodemographic factors we examined and the parent problem technology use construct.

The levels of perceived parent problematic use of technology and perceived amount by which technology interfered with daily parent-child interactions were higher in this study than in the first study reporting on these measures. In particular, in our study, parents reported an average of 3.03 (SD 2.07) devices interfering daily, compared with 2 devices in a previous study [9]. This is likely because the measurements in that study were taken between 2014 and 2016; secular trends toward increasing awareness of problematic technology use in general and technology interference in parenting in particular have occurred since then.

Limitations

The limitations of this study include those commonly associated with web-based surveys. Parents responding to web-based surveys may be more accepting of technological interventions than parents who do not respond to web-based surveys, resulting in selection bias. Furthermore, although social desirability bias may have resulted in a desire to downplay one's technology use during active parenting, previous research suggests that web-based surveys are no more susceptible to this type of bias than human-administered surveys [45].

Another limitation is that our sample had far more females (222/280, 79.2%) than are representative of the US population. This could have been because the survey focused on parenting topics. Our sample was not representative of the US population in other ways. Compared with the 2018 American Community Survey, the study's respondents reported slightly lower income levels and slightly higher education levels; fewer spoke a language other than English at home, and fewer were ethnic minority respondents [46]. In addition, we were not able to assess the degree of bias that resulted from parents who were approached but elected not to participate and the observations that had to be dropped because of data quality criteria.

Finally, the interpretation of some scales is limited by a lack of established national standards. For example, no guidelines exist for determining a sufficiently high level of technology acceptance using the TAM-PU scale.

Conclusions

AI-based tools may be acceptable to use as coaching aids to help a wide sociodemographic range of parents improve their attentiveness while caring for their young children, especially in the face of technoference from their own use of mobile devices. Designers and developmental specialists should work together to develop and test AI-based tools to reduce parenting technoference, with an initial focus on younger parents. Future investigations should validate whether it is sufficient to focus AI-based parenting supports to combat technoference on parents' general overuse of digital technology rather than their specific problems with technoference and to identify other factors that influence the acceptability and utility of these supports.

Conflicts of Interest

AM is a co-founding scientist and paid scientific advisor at Dawnlight Technologies, a company pursuing applications of AI to healthcare delivery. Other authors declare no conflicts of interest

Multimedia Appendix 1
Full questionnaire/survey.
[DOCX File , 29 KB - jmir_v23i3e19461_app1.docx]

References

1. Heckman JJ. The economics, technology, and neuroscience of human capability formation. Proc Natl Acad Sci U S A 2007 Aug 14;104(33):13250-13255 [FREE Full text] [doi: 10.1073/pnas.0701362104] [Medline: 17686985]
2. Cunha F, Heckman J. The technology of skill formation. Am Econ Rev 2007 Apr 01;97(2):31-47. [doi: 10.1257/aer.97.2.31]
3. Cunha F, Heckman JJ. The economics and psychology of inequality and human development. J Eur Econ Assoc 2009;7(2):320-364 [FREE Full text] [doi: 10.1162/jeea.2009.7.2-3.320] [Medline: 20209045]
4. Shonkoff JP, Phillips DA, editors. From neurons to neighborhoods : the science of early childhood development. Washington, D.C: National Academies Press (US); 2000.
5. Bowlby J. A secure base: parent-child attachment and healthy human development. In: Choice Reviews Online. New York: Basic Books; Apr 01, 1988.
6. Thompson RA. Social-emotional development in the first three years. The Pennsylvania State University. 2018. URL: https://www.rwjf.org/en/library/research/2018/04/social-emotional-development-in-the-first-three-years.html [accessed 2018-07-31]
7. Radesky JS, Kistin C, Eisenberg S, Gross J, Block G, Zuckerman B, et al. Parent perspectives on their mobile technology use: the excitement and exhaustion of parenting while connected. J Dev Behav Pediatr 2016;37(9):694-701. [doi: 10.1097/dbp.0000000000000357]
8. Radesky J, Leung C, Appugliese D, Miller A, Lumeng J, Rosenblum K. Maternal mental representations of the child and mobile phone use during parent-child mealtimes. J Dev Behav Pediatr 2018 May;39(4):310-317 [FREE Full text] [doi: 10.1097/DBP.0000000000000556] [Medline: 29485515]
9. McDaniel BT, Radesky JS. Technoference: Parent Distraction With Technology and Associations With Child Behavior Problems. Child Dev 2018 Jan;89(1):100-109 [FREE Full text] [doi: 10.1111/cdev.12822] [Medline: 28493400]
10. Cho K, Lee J. Influence of smartphone addiction proneness of young children on problematic behaviors and emotional intelligence: Mediating self-assessment effects of parents using smartphones. Computers in Human Behavior 2017 Jan;66:303-311. [doi: 10.1016/j.chb.2016.09.063]
11. Tsetsi E, Rains SA. Smartphone internet access and use: extending the digital divide and usage gap. Mob Med Comm 2017 Jun 13;5(3):239-255 [FREE Full text] [doi: 10.1177/2050157917708329]
12. Baker S, Sanders MR, Turner KMT, Morawska A. A randomized controlled trial evaluating a low-intensity interactive online parenting intervention, Triple P online brief, with parents of children with early onset conduct problems. Behav Res Ther 2017 Apr;91:78-90 [FREE Full text] [doi: 10.1016/j.brat.2017.01.016] [Medline: 28167330]
13. Cates CB, Weisleder A, Mendelsohn AL. Mitigating the effects of family poverty on early child development through parenting interventions in primary care. Acad Pediatr 2016 Apr;16(3 Suppl):S112-S120 [FREE Full text] [doi: 10.1016/j.acap.2015.12.015] [Medline: 27044688]
14. Hall CM, Bierman KL. Technology-assisted interventions for parents of young children: emerging practices, current research, and future directions. Early Child Res Q 2015;33:21-32 [FREE Full text] [doi: 10.1016/j.ecresq.2015.05.003] [Medline: 27773964]
15. Olds D. Improving preschool for low-income children with programmatic randomized controlled trials. Arch Pediatr Adolesc Med 2007 Aug;161(8):807-809 [FREE Full text] [doi: 10.1001/archpedi.161.8.807] [Medline: 17679664]
16. Weil LW, Middleton L. Use of the LENA tool to evaluate the effectiveness of a parent intervention program. Perspect Lang Learn Educ 2010 Oct;17(3):108-111 [FREE Full text] [doi: 10.1044/lle17.3.108]
17. Weisleder A, Fernald A. Talking to children matters: early language experience strengthens processing and builds vocabulary. Psychol Sci 2013 Nov 01;24(11):2143-2152 [FREE Full text] [doi: 10.1177/0956797613488145] [Medline: 24022649]
18. Zhang Y, Xu X, Jiang F, Gilkerson J, Xu D, Richards JA, et al. Effects of quantitative linguistic feedback to caregivers of young children: a pilot study in China. Comm Disorders Q 2015 Mar 19;37(1):16-24 [FREE Full text] [doi: 10.1177/1525740115575771]
19. Gilkerson J, Richards JA, Topping K. Evaluation of a LENA-based online intervention for parents of young children. J Early Intervention 2017 Jul 26;39(4):281-298 [FREE Full text] [doi: 10.1177/1053815117718490]
20. Mendelsohn AL, Cates CB, Weisleder A, Berkule Johnson S, Seery AM, Canfield CF, et al. Reading aloud, play, and social-emotional development. Pediatrics 2018 May;141(5):- [FREE Full text] [doi: 10.1542/peds.2017-3393] [Medline: 29632254]
21. Weisleder A, Cates CB, Dreyer BP, Berkule Johnson S, Huberman HS, Seery AM, et al. Promotion of positive parenting and prevention of socioemotional disparities. Pediatrics 2016 Feb;137(2):- [FREE Full text] [doi: 10.1542/peds.2015-3239] [Medline: 26817934]
22. Yeung S, Downing N, Fei-Fei L, Milstein A. Bedside computer vision — moving artificial intelligence from driver assistance to patient safety. N Engl J Med 2018 Apr 05;378(14):1271-1273 [FREE Full text] [doi: 10.1056/nejmp1716891]
23. Yeung S, Rinaldo F, Jopling J, Liu B, Mehra R, Downing NL, et al. A computer vision system for deep learning-based detection of patient mobilization activities in the ICU. NPJ Digit Med 2019;2:11 [FREE Full text] [doi: 10.1038/s41746-019-0087-z] [Medline: 31304360]
24. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319 [FREE Full text] [doi: 10.2307/249008]

25. Chin WW, Johnson N, Schwarz A. A fast form approach to measuring technology acceptance and other constructs. MIS Q 2008;32(4):687 [FREE Full text] [doi: 10.2307/25148867]
26. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
27. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):- [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
28. Nunes A, Limpo TF, Castro SL. Acceptance of mobile health applications: examining key determinants and moderators. Front Psychol 2019;10:2791 [FREE Full text] [doi: 10.3389/fpsyg.2019.02791] [Medline: 31920836]
29. Zhang M, Luo M, Nie R, Zhang Y. Technical attributes, health attribute, consumer attributes and their roles in adoption intention of healthcare wearable technology. Int J Med Inform 2017 Dec;108:97-109 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.09.016] [Medline: 29132639]
30. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:- [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
31. Pereira AI, Barros L. Parental cognitions and motivation to engage in psychological interventions: a systematic review. Child Psychiatry Hum Dev 2019 Jun;50(3):347-361 [FREE Full text] [doi: 10.1007/s10578-018-0852-2] [Medline: 30430390]
32. Ayers S, Baum A, McManus C, Newman S, Wallston K, Weinman J, et al. Cambridge Handbook of Psychology, Health and Medicine. 1997. URL: https://cdchester.co.uk/wp-content/uploads/2018/05/SUSANA1.pdf [accessed 2021-02-15]
33. Rosenzweig CMS. What is prime panels? Breaking down cloud research's panel management tool. URL: https://www.cloudresearch.com/resources/research/what-is-prime-panels/ [accessed 2020-11-18]
34. Chandler J, Rosenzweig C, Moss AJ, Robinson J, Litman L. Online panels in social science research: expanding sampling methods beyond Mechanical Turk. Behav Res Methods 2019 Oct;51(5):2022-2038 [FREE Full text] [doi: 10.3758/s13428-019-01273-7] [Medline: 31512174]
35. Bettiga D, Lamberti L, Lettieri E. Individuals' adoption of smart technologies for preventive health care: a structural equation modeling approach. Health Care Manag Sci 2020 Jun;23(2):203-214 [FREE Full text] [doi: 10.1007/s10729-019-09468-2] [Medline: 30684067]
36. Zhang C, Conrad F. Speeding in web surveys: the tendency to answer very fast and its association with straightlining. Surv Res Methods 2014:127-135 [FREE Full text] [doi: 10.18148/srm/2014.v8i2.5453]
37. Kim Y, Dykema J, Stevenson J, Black P, Moberg DP. Straightlining: overview of measurement, comparison of indicators, and effects in mail–web mixed-mode surveys. Soc Sci Comp Rev 2018 Feb 20;37(2):214-233 [FREE Full text] [doi: 10.1177/0894439317752406]
38. Beran TN, Violato C. Structural equation modeling in medical research: a primer. BMC Res Notes 2010 Oct 22;3:267 [FREE Full text] [doi: 10.1186/1756-0500-3-267] [Medline: 20969789]
39. Bendel RB, Afifi AA. Comparison of stopping rules in forward “Stepwise” regression. J Am Stat Asso 1977 Mar;72(357):46-53 [FREE Full text] [doi: 10.1080/01621459.1977.10479905]
40. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression, Third Edition. New Jersey: John Wiley & Sons; 2013.
41. Cohen J. Statistical power analysis for the behavioral sciences, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
42. Zhang Y, Rau PP. Playing with multiple wearable devices: Exploring the influence of display, motion and gender. Computers in Human Behavior 2015 Sep;50:148-158 [FREE Full text] [doi: 10.1016/j.chb.2015.04.004]
43. Csibi S, Griffiths MD, Demetrovics Z, Szabo A. Analysis of problematic smartphone use across different age groups within the ‘Components Model of Addiction’. Int J Ment Health Addiction 2019 May 23 [FREE Full text] [doi: 10.1007/s11469-019-00095-0]
44. van Deursen AJAM, Bolle C, Hegner SM, Kommers PA. Modeling habitual and addictive smartphone behavior. Comp Hum Beh 2015 Apr;45:411-420 [FREE Full text] [doi: 10.1016/j.chb.2014.12.039]
45. Weigold A, Weigold IK, Russell EJ. Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. Psychol Methods 2013 Mar;18(1):53-70 [FREE Full text] [doi: 10.1037/a0031607] [Medline: 23477606]
46. 2014—2018 ACS 5-Year Data Profile. URL: https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2018/ [accessed 2020-04-09]

Abbreviations

AI: artificial intelligence
HBM: Health Belief Model
LENA: Language ENvironment Analysis
PU: perceived usefulness
TAM: technology acceptance model

Edited by R Kukafka; submitted 18.04.20; peer-reviewed by S Binyamin, R Zhong; comments to author 17.08.20; revised version received 24.11.20; accepted 20.01.21; published 15.03.21.

Please cite as:
Glassman J, Humphreys K, Yeung S, Smith M, Jauregui A, Milstein A, Sanders L
Parents' Perspectives on Using Artificial Intelligence to Reduce Technology Interference During Early Childhood: Cross-sectional Online Survey
J Med Internet Res 2021;23(3):e19461
URL: https://www.jmir.org/2021/3/e19461
doi:10.2196/19461
PMID:33720026

Original Paper

Reducing the Impact of Confounding Factors on Skin Cancer Classification via Image Segmentation: Technical Model Study

Roman C Maron[1], MSc; Achim Hekler[1], MSc; Eva Krieghoff-Henning[1], PhD; Max Schmitt[1], MSc; Justin G Schlager[2], MD; Jochen S Utikal[3,4], MD; Titus J Brinker[1], MD

[1]Digital Biomarkers for Oncology Group, National Center for Tumor Diseases (NCT), German Cancer Research Center (DKFZ), Heidelberg, Germany

[2]Department of Dermatology and Allergy, University Hospital, LMU Munich, Munich, Germany

[3]Department of Dermatology, Heidelberg University, Mannheim, Germany

[4]Skin Cancer Unit, German Cancer Research Center (DKFZ), Heidelberg, Germany

Corresponding Author:
Titus J Brinker, MD
Digital Biomarkers for Oncology Group
National Center for Tumor Diseases (NCT)
German Cancer Research Center (DKFZ)
Im Neuenheimer Feld 280
Heidelberg, 69120
Germany
Phone: 49 6221 3219304
Email: titus.brinker@dkfz.de

Abstract

Background: Studies have shown that artificial intelligence achieves similar or better performance than dermatologists in specific dermoscopic image classification tasks. However, artificial intelligence is susceptible to the influence of confounding factors within images (eg, skin markings), which can lead to false diagnoses of cancerous skin lesions. Image segmentation can remove lesion-adjacent confounding factors but greatly change the image representation.

Objective: The aim of this study was to compare the performance of 2 image classification workflows where images were either segmented or left unprocessed before the subsequent training and evaluation of a binary skin lesion classifier.

Methods: Separate binary skin lesion classifiers (nevus vs melanoma) were trained and evaluated on segmented and unsegmented dermoscopic images. For a more informative result, separate classifiers were trained on 2 distinct training data sets (human against machine [HAM] and International Skin Imaging Collaboration [ISIC]). Each training run was repeated 5 times. The mean performance of the 5 runs was evaluated on a multi-source test set (n=688) consisting of a holdout and an external component.

Results: Our findings showed that when trained on HAM, the segmented classifiers showed a higher overall balanced accuracy (75.6% [SD 1.1%]) than the unsegmented classifiers (66.7% [SD 3.2%]), which was significant in 4 out of 5 runs (P<.001). The overall balanced accuracy was numerically higher for the unsegmented ISIC classifiers (78.3% [SD 1.8%]) than for the segmented ISIC classifiers (77.4% [SD 1.5%]), which was significantly different in 1 out of 5 runs (P=.004).

Conclusions: Image segmentation does not result in overall performance decrease but it causes the beneficial removal of lesion-adjacent confounding factors. Thus, it is a viable option to address the negative impact that confounding factors have on deep learning models in dermatology. However, the segmentation step might introduce new pitfalls, which require further investigations.

(J Med Internet Res 2021;23(3):e21695) doi:10.2196/21695

KEYWORDS

dermatology; diagnosis; artificial intelligence; neural networks; image segmentation; confounding factors; artifacts; melanoma; nevus; deep learning

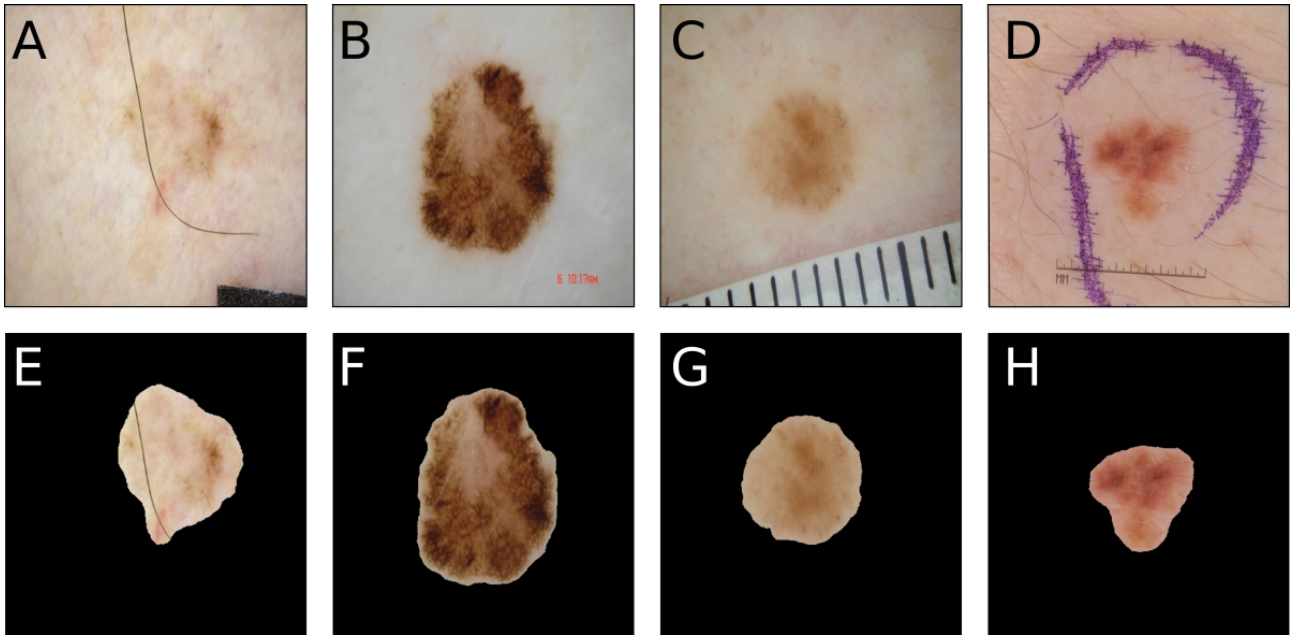
Introduction

Deep learning models have achieved impressive results in dermoscopic image skin cancer classification, as exemplified by a range of studies on binary and multiclass classification tasks [1-5]. The creation of open-source dermoscopic image databases [6-8] has enabled much of the current research in this area by facilitating the training and evaluation of deep learning models. Supervised learning is commonly used, where the deep learning model is trained on labeled training data (eg, dermoscopic image plus its corresponding diagnosis), and it continually optimizes its internal parameters. This produces an inferred function that ideally classifies previously unseen data correctly based on a valid strategy (eg, in the case of skin lesions, based on relevant biological and structural features). However, it is not uncommon for deep learning models to learn spurious correlations within the training data. As a result, these models fail when evaluating data not exhibiting the respective correlations. In image analysis, such correlations are often introduced by visual artifacts, which act as confounding factors and have been observed to result in performance degradation [9,10]. A recent dermatology study showed that skin markings significantly interfered with the correct diagnosis of nevi by deep learning convolutional neural networks (CNNs) by increasing the melanoma probability scores and consequently, the false-positive rate [11]. Besides skin markings with stains/ink, a variety of artifacts are encountered in public and proprietary dermoscopic image databases, such as dark image corners, gel bubbles, color charts, ruler marks, or skin hairs (see Figure 1).

A variety of strategies have been proposed to tackle confounding factors such as digital hair removal, image cropping, or image segmentation [12]. In image segmentation, an image is partitioned into 2 or more regions so that each region can be analyzed on its own. Dermoscopic image segmentation usually partitions the image into foreground (lesion) and background (surrounding skin, see Figure 1). This preprocessing approach has the advantage that it not only simplifies the representation of the image but also removes the surrounding artifacts. Theoretically, the image fed to the deep learning model after segmentation consists mainly of the lesion, which presumably contains the most information but the least confounding factors.

In this study, we therefore determined if and how image segmentation affects skin lesion classification performance of deep learning–based algorithms. We compared the performance of 2 workflows: one where skin lesion classifiers were trained by a traditional end-to-end approach on unsegmented dermoscopic images and one where classifiers were trained by a two-step approach on images that have undergone prior segmentation.

Figure 1. Typical artifacts encountered in dermoscopic image databases. Panels A-D show an exemplary range of artifacts often found in dermoscopic images, which are (left to right) color charts and hair, text, ruler markings, and marker ink. Panels E-H show how a corresponding segmented image could look like, with the surrounding artifacts removed but artifacts within the lesion (Panel E) still visible.

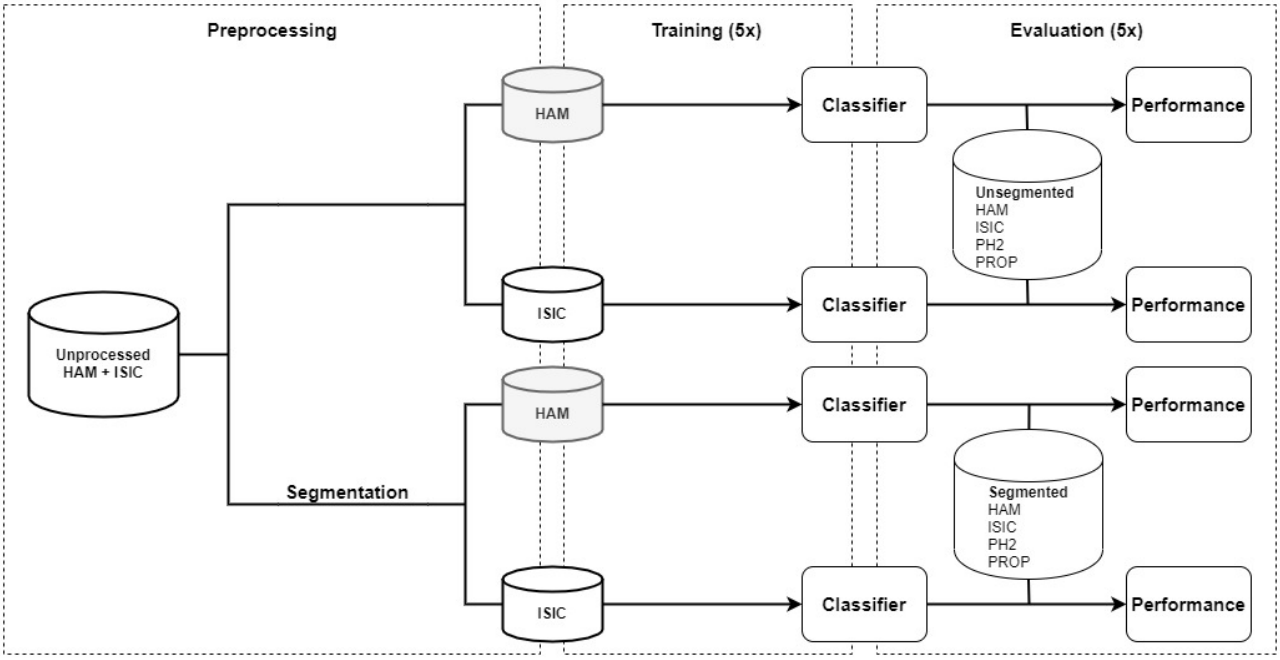


Methods

Study Design

Binary classifiers (nevus vs melanoma) were trained on 2 different data sets and on unsegmented or segmented images, respectively, resulting in 4 separate types of classifiers. All classifiers were evaluated on a test set (n=688) consisting of 1 holdout component (n=200) and 3 external components (n=488). For each classifier type, 5 training and testing runs were performed in order to obtain robust performance estimates, which encompass the stochastic nature of the training process (see Figure 2). Ethics approval was waived by the ethics committee of the University of Heidelberg, as images were open source and anonymous.

Figure 2. Flowchart of the study design. A training data set consisting of images from 2 different sources was either segmented or not segmented and split into 2 smaller partitions based on image origin (HAM or ISIC). An individual classifier was then trained on each of the 4 training sets and evaluated on a multi-source test set, which underwent a preprocessing step that equaled the training data preprocessing. Training and evaluation were repeated a total of 5 times for a more robust measure. HAM: human against machine data set; ISIC: international skin imaging collaboration data set; PH2: hospital Pedro Hispano data set; PROP: proprietary data set.



Data Sets

Dermoscopic images for developing the segmentation model were obtained from task 1 of the International Skin Imaging Collaboration (ISIC) 2018 challenge [7,13]. This data set is already split into a training, validation, and test set by the challenge organizers and contains dermoscopic lesion images together with a binary image mask, which partitions the image into a background (areas outside the primary lesion) and foreground (areas inside the primary lesion). This mask represents the "ground truth" with respect to the correct partitioning of the images. Dermoscopic images for developing the skin lesion classifiers were obtained from 2 sources: from part 3 of the ISIC 2017 challenge [6] and from the human against machine (HAM) data set [7]. Both data sets are mutually exclusive, with the HAM data set showing considerably fewer artifacts than the ISIC data set. Duplicated images within the HAM data set were removed prior to splitting the data set into the training, validation, and test set. The ISIC 2017 challenge data set had already been split by the challenge organizers.

Two additional external data sets were used for classifier evaluation. The first data set is publicly available and contains dermoscopic images acquired at the Dermatology Service of Hospital Pedro Hispano (PH2), Matosinhos, Portugal [8]. The second data set is proprietary (PROP) and contains dermoscopic images acquired from the Department of Dermatology and Allergy, University Hospital, LMU Munich, Munich and from the Department of Dermatology, Heidelberg University, Mannheim. Both data sets also contain some of the artifacts observed in ISIC and HAM, such as black image corners, rulers, or skin markings. As PH2 also contains binary image masks from dermatologists, this data set was also used for the evaluation of the segmentation model. Details on the training, validation, and test set composition are listed in Table S1 of Multimedia Appendix 1.

Segmentation Model and Classifier Development

For image segmentation, a CNN in the form of a U-Net was employed [14]. The model's raw output, which consists of a binary image mask, was further automatically processed by removing noise, closing holes, and replacing empty masks. Skin lesion classifiers were generated using a ResNet50 architecture, which was pretrained on ImageNet. For details on segmentation model and classifier development, refer to the supplementary methods (Multimedia Appendix 1).

Analysis

The segmentation model's performance was evaluated using a thresholded mean Jaccard index, a score between 0 and 1, which measures the similarity between the ground truth mask and the model's output mask. The threshold was based on the ISIC 2018 challenge and set to 0.65, meaning that any lower scores were set to 0. The performance for each individual classifier was measured using balanced accuracy as the primary endpoint, with sensitivity, specificity, and area under the receiver operating characteristic curve (AUROC) as secondary endpoints. As we repeated each classifier training and evaluation step 5 times, metrics were first computed for each individual classifier and then averaged to obtain a mean performance measure. Performance comparisons were carried out between the preprocessing methods (ie, segmented vs unsegmented) and not between the underlying training data sets (ie, not HAM vs ISIC). Thus, we compared HAM segmented to HAM unsegmented and ISIC segmented to ISIC unsegmented but not HAM segmented to ISIC segmented. Statistical significance was evaluated for the primary endpoint by using a two-sided McNemar test and considered significant at $P<.005$ (Bonferroni

correction by a factor of 10) to account for multiple testing when comparing the individual segmented HAM/ISIC classifiers to unsegmented HAM/ISIC classifiers for each of the 5 runs (one-on-one comparison). *P* values are listed only for significant runs.

Results

Segmentation Model Performance

The thresholded Jaccard index on the ISIC holdout test set for the segmentation model after mask postprocessing was 0.75 and increased to 0.81 on the external PH2 set.

Classifier Performance

The overall balanced accuracy was numerically higher for the unsegmented ISIC classifiers (78.3% [SD 1.8%]) than for the segmented ISIC classifiers (77.4% [SD 1.5%]). This was significantly different in 1 out of 5 runs (P=.004). When trained on HAM, the segmented classifiers showed a higher overall balanced accuracy (75.6% [SD 1.1%]) than the unsegmented classifiers (66.7% [SD 3.2%]). This difference was significant for 4 out of the 5 classifiers (P<.001). A subanalysis of the performance on the holdout and external test set component shows that segmented classifiers had a numerically higher overall balanced accuracy on the external component than unsegmented classifiers, regardless of the data set source (see Table 1). The reverse trend was observed for the holdout component. AUROC followed the same trends as mean balanced accuracy.

Table 1. Overview of the balanced accuracy and area under the receiver operating characteristic curve for each type of classifier across the holdout, external, and overall test set.

Test set components, metric	Trained classifiers			
	HAM[a] segmented (%)	HAM unsegmented (%)	ISIC[b] segmented (%)	ISIC unsegmented (%)
Holdout				
Balanced accuracy, mean (SD)	87.6 (1.4)	*89.4 (0.9)* [c]	77.1 (1.5)	*80.0 (2.6)*
AUROC[d], mean (SD)	0.95 (0.006)	*0.964 (0.002)*	0.839 (0.008)	*0.89 (0.1)*
External				
Balanced accuracy, mean (SD)	*69.9 (1.3)*	57.6 (4.1)	*78.2 (1.6)*	77.6 (1.7)
AUROC, mean (SD)	*0.765 (0.011)*	0.647 (0.025)	*0.874 (0.005)*	0.851 (0.018)
Overall				
Balanced accuracy, mean (SD)	*75.6 (1.1)*	66.7 (3.2)	77.4 (1.5)	*78.3 (1.8)*
AUROC, mean (SD)	*0.841 (0.008)*	0.763 (0.02)	0.856 (0.005)	*0.862 (0.014)*

[a]HAM: human against machine data set.

[b]ISIC: International Skin Imaging Collaboration data set.

[c]The italicized data indicate the higher metric when comparing between classifiers trained on a segmented/unsegmented version of the same data set.

[d]AUROC: area under the receiver operating characteristic curve.

ISIC classifiers (regardless of preprocessing) show a comparable balanced accuracy across the holdout and external test set components, resulting in a similar balanced accuracy for the overall test set. In contrast, the segmented HAM classifiers show a substantially higher overall balanced accuracy to the unsegmented HAM classifiers. This better overall balanced accuracy stems from a visible performance difference on the external test set component, which is largely driven by a drop in the balanced accuracy for PH2. Here, the balanced accuracy of unsegmented HAM classifiers was 63.2% (SD 7.1%) compared to 84.4% (SD 2.9%) for the segmented HAM classifiers (see Table 2). Equivalent tables showing the results for the metric sensitivity and specificity are found in Table S2 and Table S3 of Multimedia Appendix 1.

Table 2. Overview of the balanced accuracy and area under the receiver operating characteristic curve for each type of classifier across the external test set's 3 individual components.

External test set components, metric	Trained classifiers			
	HAM[a] segmented (%)	HAM unsegmented (%)	ISIC[b] segmented (%)	ISIC unsegmented (%)
HAM/ISIC[c]				
Balanced accuracy, mean (SD)	*61.0 (1.3)* [d]	58.9 (3.1)	74.1 (3.6)	*76.5 (1.8)*
AUROC[e], mean (SD)	0.628 (0.005)	*0.636 (0.023)*	0.827 (0.019)	*0.851 (0.022)*
PH2[f]				
Balanced accuracy, mean (SD)	*84.4 (2.9)*	63.2 (7.1)	*86.4 (1.3)*	83.7 (0.8)
AUROC, mean (SD)	*0.928 (0.022)*	0.894 (0.021)	*0.947 (0.007)*	0.912 (0.018)
PROP[g]				
Balanced accuracy, mean (SD)	71.1 (1.8)	*75.7 (4.2)*	68.7 (1.6)	*74.6 (2.8)*
AUROC, mean (SD)	0.825 (0.033)	*0.857 (0.034)*	*0.88 (0.025)*	0.814 (0.015)

[a]HAM: human against machine data set.

[b]ISIC: International Skin Imaging Collaboration data set.

[c]If classifiers were trained on HAM images, the first external test set component consists of ISIC and vice versa.

[d]The italicized data indicate the higher metric when comparing between classifiers trained on a segmented/unsegmented version of the same data set.

[e]AUROC: area under the receiver operating characteristic curve.

[f]PH2: hospital Pedro Hispano data set.

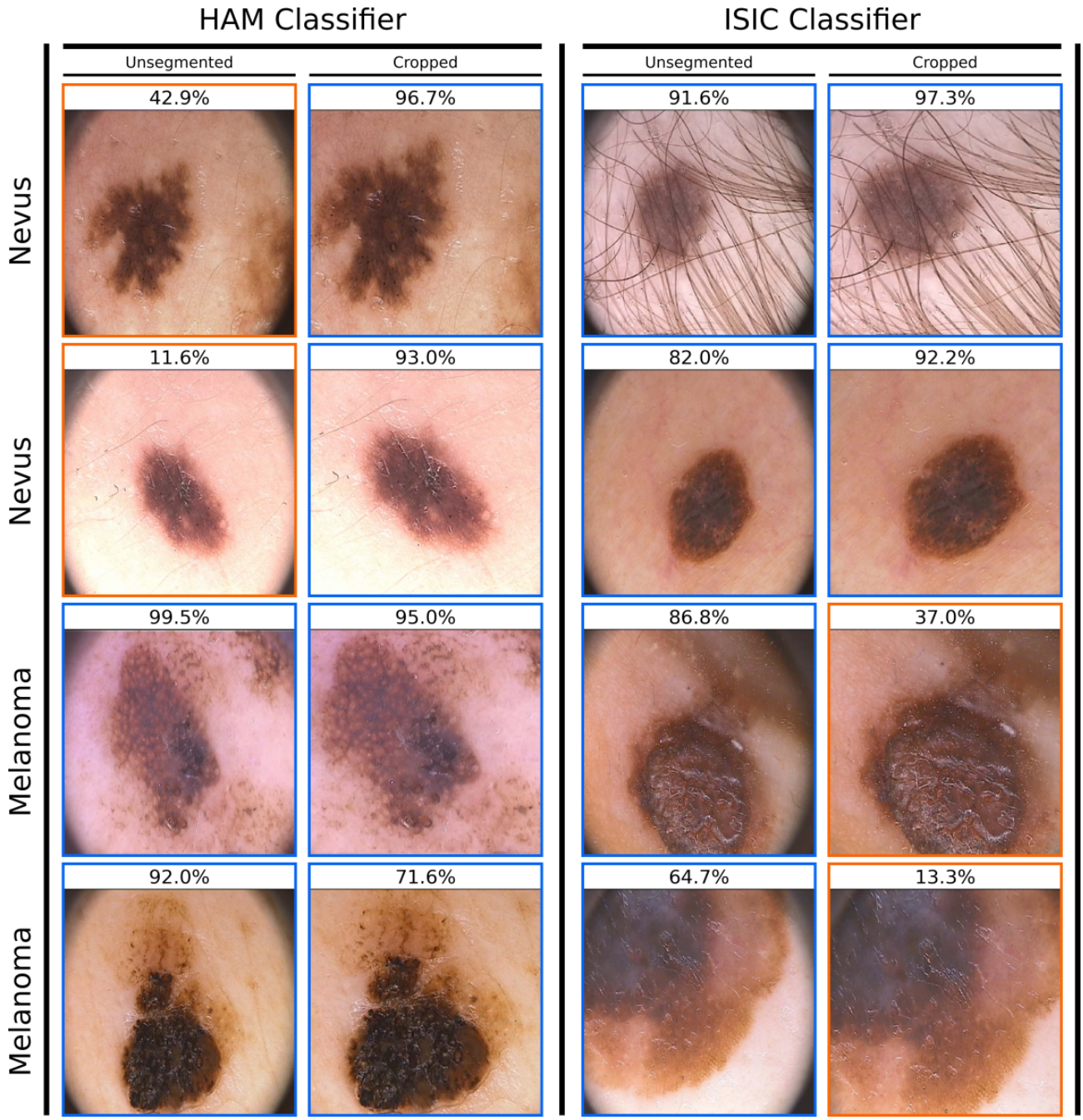
[g]PROP: proprietary data set.

Additional Analyses

Some additional analyses were carried out based on the obtained results. As the unsegmented HAM classifiers showed poor performance on PH2 with high sensitivity (95.5% [SD 1.9%]) but low specificity (30.9% [SD 13.6%], Table S3 of Multimedia Appendix 1), their performance was again evaluated on cropped unsegmented PH2 images. As the PH2 data set consists of images with predominantly black corners (see Figure 3), we speculated that these could be artifacts, which caused the drop in performance. We therefore manually cropped all unsegmented PH2 images just enough so that any black corner was removed. On cropped PH2 images, specificity increased to 65.8% (SD 8.3%) at almost unchanged sensitivity of 93.5% (SD 3%), resulting in an overall mean balanced accuracy of 79.6% (SD 3.8%). As the unsegmented ISIC classifiers showed a comparable performance to the segmented ISIC classifiers, there was no reason to assume that these classifiers are also negatively influenced by black image corners. However, when its performance was evaluated on cropped PH2, sensitivity decreased from 82% (SD 2.9%) (unsegmented) to 67.5% (SD 5.7%) (cropped) with specificity increasing from 85.4% (SD 2.8%) to 89.4% (SD 1.7%), resulting in a change of mean balanced accuracy from 83.7% (SD 0.8%) to 78.4% (SD 3.6%).

As ground truth segmentation masks were available for the PH2 data set, PH2 images were experimentally segmented using these masks instead of the masks produced by the segmentation model and subsequently used for evaluation. These masks were produced by an expert dermatologist; therefore, a similar performance was expected. However, segmented HAM and ISIC classifiers showed a lower balanced accuracy for PH2 images when processed by the ground truth masks (82% [SD 2.2%] and 76.1% [SD 2.9%], respectively) as opposed to the segmentation model mask (84.4% [SD 2.9%] and 86.4% [SD 1.3%], respectively). This change resulted from a drop in specificity from 80.2% (SD 3.8%) and 82.4% (SD 5.0%) to 76.5% (SD 5.6%) and 63.2% (SD 9.7%) at almost constant sensitivity (88.5% [SD 5.1%] and 90.5% [SD 3.7%] vs 87.5% [SD 5.0%] and 89.0% [SD 5.1%], respectively).

Figure 3. Exemplary predictions of a classifier trained on unsegmented HAM (left) and ISIC (right) images and evaluated on unsegmented and cropped PH2 images. The target class (ground truth) for each lesion is displayed to the left, with the classifier's output probability for the target class on top. An output probability larger than 50% corresponds to a correct classification, which is also indicated by a blue frame, whereas an orange frame denotes an incorrect classification. HAM: human against machine data set; ISIC: international skin imaging collaboration data set; PH2: hospital Pedro Hispano data set.



Discussion

Overview of the Study

In this study, we established and compared the performance of 2 classification workflows. The first workflow did not include a preprocessing step, and training and test set images were unmodified. The second included preprocessing where images were segmented prior to classifier training and evaluation. For training, we used 2 distinct training data sets (HAM and ISIC) and established the performance on a multi-source test set. Our findings show that while performance is highly dependent on the source of the training and test set, segmentation does not lead to an overall decrease in the performance of a ResNet50 architecture and may even lead to an improved classifier, which is presumably at a decreased risk to suffer from common lesion-adjacent confounding factors.

Principal Results

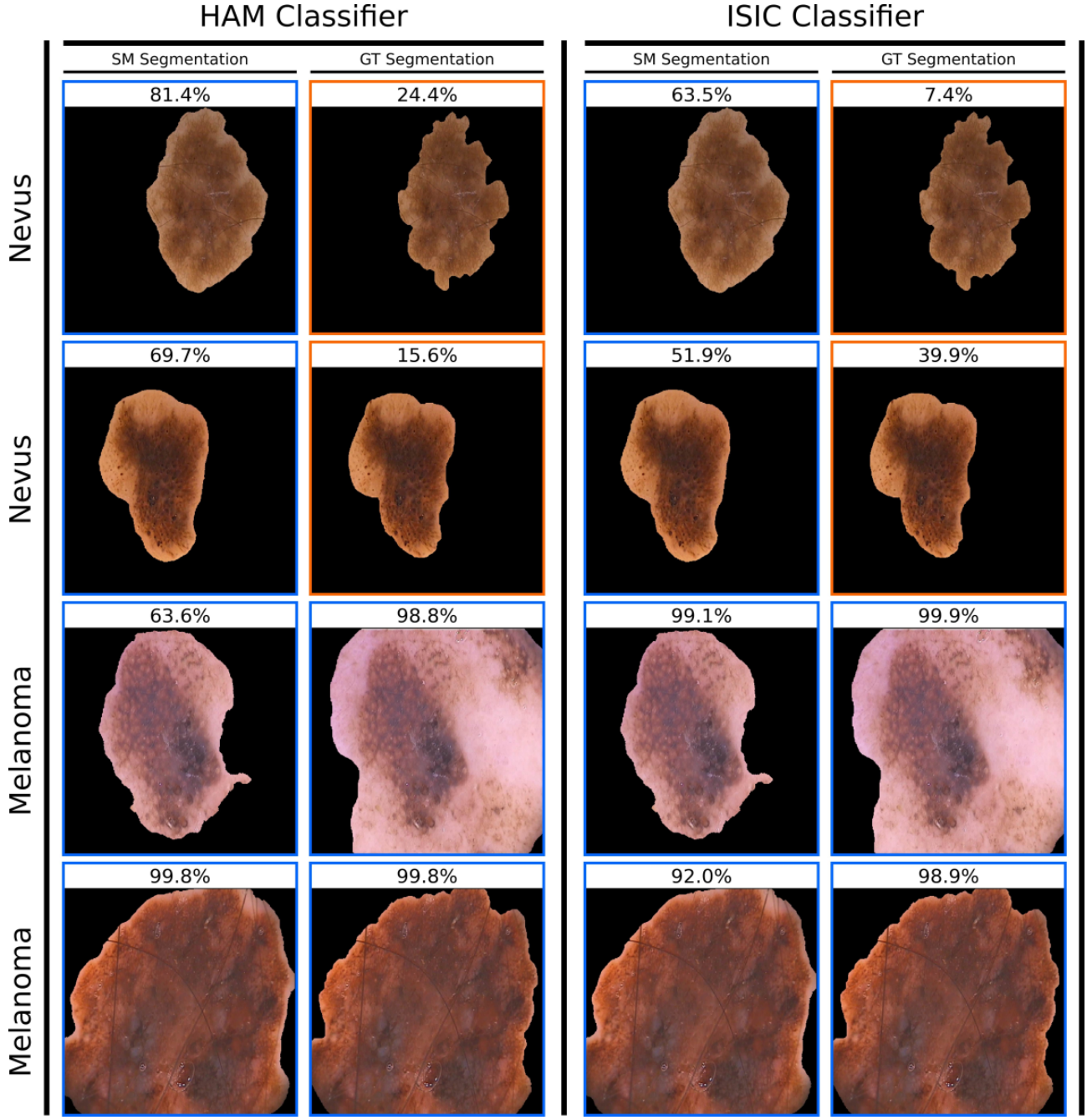
The overall comparable performance of classifiers trained on segmented and unsegmented images shows that classifiers are able to distinguish melanoma from nevus images based largely on the lesion itself without requiring the surrounding skin area for additional information. This is not unexpected as visual inspection by dermatologists also mainly focuses on global and local features within the lesion, although features such as

increased vascularization of the surrounding skin and lesions on sun-damaged or aged skin are associated with a higher risk of skin cancer and can thus be used as cues. While segmentation requires an extra step compared to end-to-end classification, it may be worthwhile as proper segmentation removes potential preexisting confounding factors surrounding the lesion (albeit not within the lesion, eg, hairs, overlapping rulers). Given the prevalence and large variety of artifacts in public dermoscopic databases such as ISIC, such measures are warranted to counteract the possibility of the classifier incorporating confounding factors in its decision process. For example, gentian violet skin markers were previously shown to be associated with a higher melanoma probability by a CNN approved for use as a medical device in the European market [11]. As artifact perception by a CNN-based classifier is dependent on the constitution of the underlying training data, this finding is not necessarily applicable to other CNN-based classifiers but highlights the negative impact of artifacts that may manifest themselves in a variety of ways. In this study, we hypothesize that classifiers trained on unsegmented HAM and ISIC images correlate black image corners with the occurrence of melanoma, albeit to varying degrees. Both unsegmented classifiers were evaluated on unsegmented and cropped PH2 images, where cropping completely removed the black corners (see Figure 3). In both cases, specificity increased when using cropped PH2 images. Sensitivity remained almost unchanged for the HAM classifiers and decreased for the ISIC classifiers, suggesting that classifiers trained on either training set associate black image corners with melanoma, but weigh its importance differently. Alternatively, it cannot be ruled out that the observed performance change stems from the cropping process, which introduces resolution changes, image distortions, and the removal of potentially relevant biological information if parts of the lesion are cropped out. However, given the one-sided performance increase (ie, for specificity) for classifiers from both data sets and the large prevalence of black image corners in the HAM and ISIC training data, a correlation is not unlikely.

As the segmentation step lies upstream of the classifier training and evaluation steps, the latter two are highly dependent on the output quality of the former. While the model employed in this study achieved a threshold Jaccard index lower than the score obtained by the ISIC 2018 challenge winners (0.75 vs 0.80), a general visual inspection of the segmentation masks suggested sufficient quality (ie, lesions visible with large portions of the background adequately removed). Further evaluation of its performance on an external data set (PH2) indicated that the segmentation model generalizes well and can be employed for segmenting images from external data sets. Assuming that classifier performance is partially indicative of segmentation performance, the segmentation model generalized adequately for HAM and PH2 images (known of course for the latter already due to the ground truth masks, but confirmed here again). In contrast, classifiers trained on segmented images performed worse on PROP with low mean balanced accuracies. Given that classifiers trained on unsegmented images did not suffer from this issue, insufficient segmentation masks are a possible candidate for the problem. This is, however, difficult to verify due to the nonexisting ground truth masks. This illustrates that the performance of segmented classifiers is ultimately tied to the performance of the segmentation model.

In practice, identifying and fixing obviously faulty segmentation masks manually at test time should be feasible but may not be sufficient. As seen for the analysis of the PH2 set, where model segmentation masks were compared to ground truth segmentation masks, classifier performance may be strongly influenced by the precise way that the segmentation is done, with small differences causing large negative effects (see Figure 4). Training sets of automatically segmented images could contain their own kind of artifacts introduced by the automated segmentation process. We speculate that masks produced by the segmentation model have distinctive visual characteristics based on its training set and postprocessing methods. For instance, a certain amount of the adjacent skin may be included or the segmentation creates unique borders (eg, smoothness of edges). Any classifier trained on images with such segmentation masks might pick up on such subtleties and become susceptible to segmentation masks of a different variety.

Figure 4. Exemplary predictions of a classifier trained on unsegmented HAM (left) and ISIC (right) images and evaluated on PH2 images with different segmentation masks. PH2 images in the SM column were segmented using the segmentation model. PH2 images in the GT column were segmented using dermatologist-generated ground truth segmentation masks. The target class (ground truth) for each lesion is displayed to the left, with the classifier's output probability for the target class on top. An output probability larger than 50% corresponds to a correct classification, which is also indicated by a blue frame, whereas an orange frame denotes an incorrect classification. GT: ground truth; HAM: human against machine data set; ISIC: international skin imaging collaboration data set; PH2: hospital Pedro Hispano data set; SM: segmentation model.



Future Work and Limitations

While the study aimed at only comparing the performance of 2 classification workflows where classifiers were trained on segmented/unsegmented images, there is notable performance variation dependent on the training and test sets. While the HAM training set contained more unique melanoma lesions (514 vs 374, Table S1 of Multimedia Appendix 1), the ISIC training data set contained more images of biopsy-verified lesions and thus, probably more borderline cases. These distinct features may be advantageous or detrimental for classifier performance on any given test set. Future work should address the issue of faulty segmentation masks and closely investigate the potential artifacts arising from an upstream segmentation step. As classification was done in a binary instead of a multi-class setting due to limited data availability, these findings might not generalize to a multi-class setting. Furthermore, performance was only shown here for 1 architecture; thus, generalizability to similar architectures, while expected, is not guaranteed.

Conclusion

Skin lesion classifiers trained and evaluated on segmented images have an overall comparable performance to classifiers

trained and evaluated on unsegmented images that show the exact same lesion. In addition, segmentation comes with the added benefit of removing lesion-adjacent artifacts, which may act as confounding factors. However, this benefit comes at a cost, as classifier performance is tied to the segmentation quality. Further, image segmentation may introduce new pitfalls. Hence, further investigation is required to elucidate the effects of segmentation observed in this study.

Acknowledgments

This study was funded by the Federal Ministry of Health, Berlin, Germany (grant: Skin Classification Project; grant holder: Titus J. Brinker, MD, German Cancer Research Center, Heidelberg, Germany). The sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Conflicts of Interest

TJB reports owning a company that develops mobile apps (Smart Health Heidelberg GmbH, Handschuhsheimer Landstr. 9/1, 69120 Heidelberg). JSU is on the advisory board or has received honoraria and travel support from Amgen, Bristol Myers Squibb, GSK, LeoPharma, Merck Sharp and Dohme, Novartis, Pierre Fabre, Roche, outside the submitted work. No other disclosures were reported.

Multimedia Appendix 1
Supplementary content.
[DOCX File , 15 KB - jmir_v23i3e21695_app1.docx]

References

1. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Jan 25;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
2. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Reader study level-I and level-II Groups, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol 2018 Aug 01;29(8):1836-1842 [FREE Full text] [doi: 10.1093/annonc/mdy166] [Medline: 29846502]
3. Tschandl P, Rosendahl C, Akay BN, Argenziano G, Blum A, Braun RP, et al. Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks. JAMA Dermatol 2019 Jan 01;155(1):58-65 [FREE Full text] [doi: 10.1001/jamadermatol.2018.4378] [Medline: 30484822]
4. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, et al. Deep neural networks are superior to dermatologists in melanoma image classification. Eur J Cancer 2019 Sep;119:11-17 [FREE Full text] [doi: 10.1016/j.ejca.2019.05.023] [Medline: 31401469]
5. Maron RC, Weichenthal M, Utikal JS, Hekler A, Berking C, Hauschild A, Collabrators. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. Eur J Cancer 2019 Sep;119:57-65 [FREE Full text] [doi: 10.1016/j.ejca.2019.06.013] [Medline: 31419752]
6. Codella N, Gutman D, Emre CM, Helba B, Marchetti M, Dusza S, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). 2018 Apr 04 Presented at: IEEE 15th International Symposium on Biomedical Imaging (ISBI); 2018; Washington, DC, USA p. 168-172. [doi: 10.1109/ISBI.2018.8363547]
7. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 2018 Aug 14;5:180161 [FREE Full text] [doi: 10.1038/sdata.2018.161] [Medline: 30106392]
8. Mendonca T, Ferreira P, Marques J, Marcal A, Rozeira J. PH² - a dermoscopic image database for research and benchmarking. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:5437-5440. [doi: 10.1109/EMBC.2013.6610779] [Medline: 24110966]
9. Lapuschkin S, Wäldchen S, Binder A, Montavon G, Samek W, Müller KR. Unmasking Clever Hans predictors and assessing what machines really learn. Nat Commun 2019 Mar 11;10(1):1096 [FREE Full text] [doi: 10.1038/s41467-019-08987-4] [Medline: 30858366]
10. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. PLoS Med 2018 Nov;15(11):e1002683 [FREE Full text] [doi: 10.1371/journal.pmed.1002683] [Medline: 30399157]
11. Winkler JK, Fink C, Toberer F, Enk A, Deinlein T, Hofmann-Wellenhof R, et al. Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition. JAMA Dermatol 2019 Aug 14:1135 [FREE Full text] [doi: 10.1001/jamadermatol.2019.1735] [Medline: 31411641]

12. Okur E, Turkan M. A survey on automated melanoma detection. Engineering Applications of Artificial Intelligence 2018 Aug;73:50-67. [doi: 10.1016/j.engappai.2018.04.028]
13. Codella N, Rotemberg V, Tschandl P, Emre CM, Dusza S, Gutman D, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the International Skin Imaging Collaboration (ISIC). arXiv csCV. URL: http://arxiv.org/abs/1902.03368 [accessed 2021-03-11]
14. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015 Oct 05 Presented at: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015; 2015; Munich, Germany p. 234-241. [doi: 10.1007/978-3-319-24574-4_28]

Abbreviations

AUROC: area under the receiver operating characteristic curve
CNN: convolutional neural network
HAM: human against machine
ISIC: International Skin Imaging Collaboration
PH2: hospital Pedro Hispano
PROP: proprietary data set

Edited by G Eysenbach; submitted 22.06.20; peer-reviewed by CT Cheng, E Frontoni, J Robinson; comments to author 28.10.20; revised version received 13.11.20; accepted 08.02.21; published 25.03.21.

Please cite as:
Maron RC, Hekler A, Krieghoff-Henning E, Schmitt M, Schlager JG, Utikal JS, Brinker TJ
Reducing the Impact of Confounding Factors on Skin Cancer Classification via Image Segmentation: Technical Model Study
J Med Internet Res 2021;23(3):e21695
URL: https://www.jmir.org/2021/3/e21695
doi:10.2196/21695
PMID:33764307

Original Paper

A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study

Judith J Prochaska[1], MPH, PhD; Erin A Vogel[1], PhD; Amy Chieng[1], BA; Matthew Kendra[2], PhD; Michael Baiocchi[3], PhD; Sarah Pajarito[4], MA; Athena Robinson[4], PhD

[1]Stanford Prevention Research Center, School of Medicine, Stanford University, Stanford, CA, United States

[2]Department of Psychiatry & Behavioral Sciences, School of Medicine, Stanford University, Stanford, CA, United States

[3]Department of Epidemiology & Population Health, School of Medicine, Stanford University, Stanford, CA, United States

[4]Woebot Health, San Francisco, CA, United States

Corresponding Author:
Judith J Prochaska, MPH, PhD
Stanford Prevention Research Center
School of Medicine
Stanford University
Medical School Office Building, X316
1265 Welch Road
Stanford, CA, 94305
United States
Phone: 1 650 724 3608
Email: jpro@stanford.edu

Abstract

Background: Misuse of substances is common, can be serious and costly to society, and often goes untreated due to barriers to accessing care. Woebot is a mental health digital solution informed by cognitive behavioral therapy and built upon an artificial intelligence–driven platform to deliver tailored content to users. In a previous 2-week randomized controlled trial, Woebot alleviated depressive symptoms.

Objective: This study aims to adapt Woebot for the treatment of substance use disorders (W-SUDs) and examine its feasibility, acceptability, and preliminary efficacy.

Methods: American adults (aged 18-65 years) who screened positive for substance misuse without major health contraindications were recruited from online sources and flyers and enrolled between March 27 and May 6, 2020. In a single-group pre/postdesign, all participants received W-SUDs for 8 weeks. W-SUDs provided mood, craving, and pain tracking and modules (psychoeducational lessons and psychotherapeutic tools) using elements of dialectical behavior therapy and motivational interviewing. Paired samples t tests and McNemar nonparametric tests were used to examine within-subject changes from pre- to posttreatment on measures of substance use, confidence, cravings, mood, and pain.

Results: The sample (N=101) had a mean age of 36.8 years (SD 10.0), and 75.2% (76/101) of the participants were female, 78.2% (79/101) were non-Hispanic White, and 72.3% (73/101) were employed. Participants' W-SUDs use averaged 15.7 (SD 14.2) days, 12.1 (SD 8.3) modules, and 600.7 (SD 556.5) sent messages. About 94% (562/598) of all completed psychoeducational lessons were rated positively. From treatment start to end, in-app craving ratings were reduced by half (87/101, 86.1% reporting cravings in the app; odds ratio 0.48, 95% CI 0.32-0.73). Posttreatment assessment completion was 50.5% (51/101), with better retention among those who initially screened higher on substance misuse. From pre- to posttreatment, confidence to resist urges to use substances significantly increased (mean score change +16.9, SD 21.4; P<.001), whereas past month substance use occasions (mean change −9.3, SD 14.1; P<.001) and scores on the Alcohol Use Disorders Identification Test-Concise (mean change −1.3, SD 2.6; P<.001), 10-item Drug Abuse Screening Test (mean change −1.2, SD 2.0; P<.001), Patient Health Questionnaire-8 item (mean change 2.1, SD 5.2; P=.005), Generalized Anxiety Disorder-7 (mean change −2.3, SD 4.7; P=.001), and cravings scale (68.6% vs 47.1% moderate to extreme; P=.01) significantly decreased. Most participants would recommend W-SUDs to a friend (39/51, 76%) and reported receiving the service they desired (41/51, 80%). Fewer felt W-SUDs met most or all of their needs (22/51, 43%).

Conclusions: W-SUDs was feasible to deliver, engaging, and acceptable and was associated with significant improvements in substance use, confidence, cravings, depression, and anxiety. Study attrition was high. Future research will evaluate W-SUDs in a randomized controlled trial with a more diverse sample and with the use of greater study retention strategies.

Trial Registration: ClinicalTrials.gov NCT04096001; http://clinicaltrials.gov/ct2/show/NCT04096001.

(J Med Internet Res 2021;23(3):e24850) doi:10.2196/24850

KEYWORDS

artificial intelligence; conversational agent; chatbot; addiction; substance misuse; treatment; acceptability; feasibility; craving; psychoeducation; psychotherapeutic; mobile phone

Introduction

Misuse of substances is common, can be serious and costly to society, and often goes untreated due to barriers to accessing care. Globally, 3.5 million people die from alcohol and illicit drug use each year [1]. The disease burden of alcohol and illicit drug addiction is the highest in the United States [2]. Over 20 million Americans (aged 12 years and older) had a substance use disorder (SUD) in 2018, 73% had an alcohol use disorder, 40% had an illicit drug use disorder, and 13% had both alcohol and illicit drug use disorders [3]. Approximately half (47%) of Americans with an SUD had a co-occurring mental illness. Treatment of depression and anxiety, the most common psychiatric comorbidities among patients with SUDs, may reduce craving and substance use and enhance overall outcomes [4].

In 2018, less than 1 in 5 individuals with a SUD received addiction treatment [3]. Alcohol and illicit drug misuse and addiction cost the United States over US $440 billion annually in lost workplace productivity, health care expenses, and crime-related costs [5]. Potential effects on individuals include an array of physical and mental health problems, overdose, trauma, and violence [5].

Web-based interventions and digital health apps may reduce or eliminate common, significant barriers to traditional SUD treatment (eg, stigma; financial, time, and transportation constraints; lack of access to qualified providers; challenges navigating complex treatment systems; and low perceived utility) [6]. Preliminary evidence suggests that digital SUD interventions affect substance use behavior [6,7] and have the potential to reduce the population burden of SUDs. To date, most digital SUD interventions have been delivered on a web platform, rather than via mobile apps. The widespread use of smartphones makes app-based intervention delivery a viable and scalable medium. In 2019, about 8 out of 10 White, Black, and Latinx adults owned a smartphone [8]. Although lower-income adults were less likely to own a smartphone than higher-income adults, they were more likely to rely on smartphones for internet access [9]. In a 2015 survey, 58% of mobile phone owners reported downloading a health app [10]. Texting is the most widely and frequently used app on a smartphone, with 97% of Americans texting at least once a day [11].

Automated conversational agents can deliver a coach-like or sponsor-like experience and yet do not require human implementation assistance for in-the-moment treatment delivery. As recent meta-analytic work suggests, conversational text-based agents may increase engagement and enjoyment in digitized mental health care [12], whereas most general mental health care apps face difficulty sustaining engagement with high dropout [13,14]. Conversational agents can provide real-time support to address substance use urges, unlike traditional in-person frameworks of weekly visits. The scale potential of conversational agents is unconstrained, immediate, and available to users in an instant [12]. Being nonhuman based also reduces perceived stigma. A study found that people were significantly more likely to disclose personal information to artificial intelligence when they believed it was computer- rather than human-monitored [15]. Users can develop a strong therapeutic alliance in the absence of face-to-face contact [16], even with a nonhuman app [17]. Digital environments can promote honest disclosure due to greater ease of processing thoughts [16] and reduced risk of embarrassment [17]. Finally, although conversational agents can present in different modalities, including text, verbal [18,19], and animation [20-25], preliminary research on modality for psychoeducation delivery specifically found that text-based presentation resulted in higher program adherence than verbal presentation [26].

Evidence for conversational agent interventions for addressing mental health problems is growing quickly and appears promising with regard to acceptability and efficacy [27]. Developed as a mental health digital app, Woebot is a text-based conversational agent available to check in with users whenever they have smartphone access. Using conversational tones, Woebot is designed to encourage mood tracking and to deliver general psychoeducation as well as tailored empathy, cognitive behavioral therapy (CBT)–based behavior change tools, and behavioral pattern insight. Among a sample of adults (N=70) randomly assigned to Woebot or an information only control group, Woebot users had statistically and clinically significant reductions in depressive symptoms ($F_{1,48}$=6.03; P=.02) after 2 weeks of use, whereas those in the control group did not. Engagement with the app was high (averaging 12 interactions within 14 days) [18].

However, the efficacy of conversational agents for treating SUDs remains unknown. Woebot's app-based platform and user-centered design philosophy make it a promising modality for SUD treatment delivery; it offers immediate, evidence-based tailored support in the peak moment of craving. An informal poll of Woebot users (in July 2018) indicated that 63% had interest in content addressing SUDs; 22% of surveyed users reported having 5 or more alcoholic drinks in a row within a

couple of hours (ie, binge use) [28], and 5% endorsed using nonprescription drugs.

Although the efficacy of automated conversational agent digital therapeutics for SUDs is still untested, such products are commercially available, and few consumers are aware that the products lack evidence [29]. This study aims to adapt the original Woebot for the treatment of SUDs (W-SUDs), and test the feasibility, acceptability, and preliminary efficacy in a single-group pre-/posttreatment design.

Methods

Study Design

In a single-group design, we examined within-subject changes in self-reported substance use behavior, cravings, confidence to resist urges to use substances, mood symptoms (depression, anxiety), and pain from pre- to posttreatment. Intervention engagement data were collected from the Woebot app during the 8-week treatment period. Acceptability ratings were collected within the app and within the posttreatment survey. The study procedures were approved by the Institutional Review Board of Stanford Medicine.

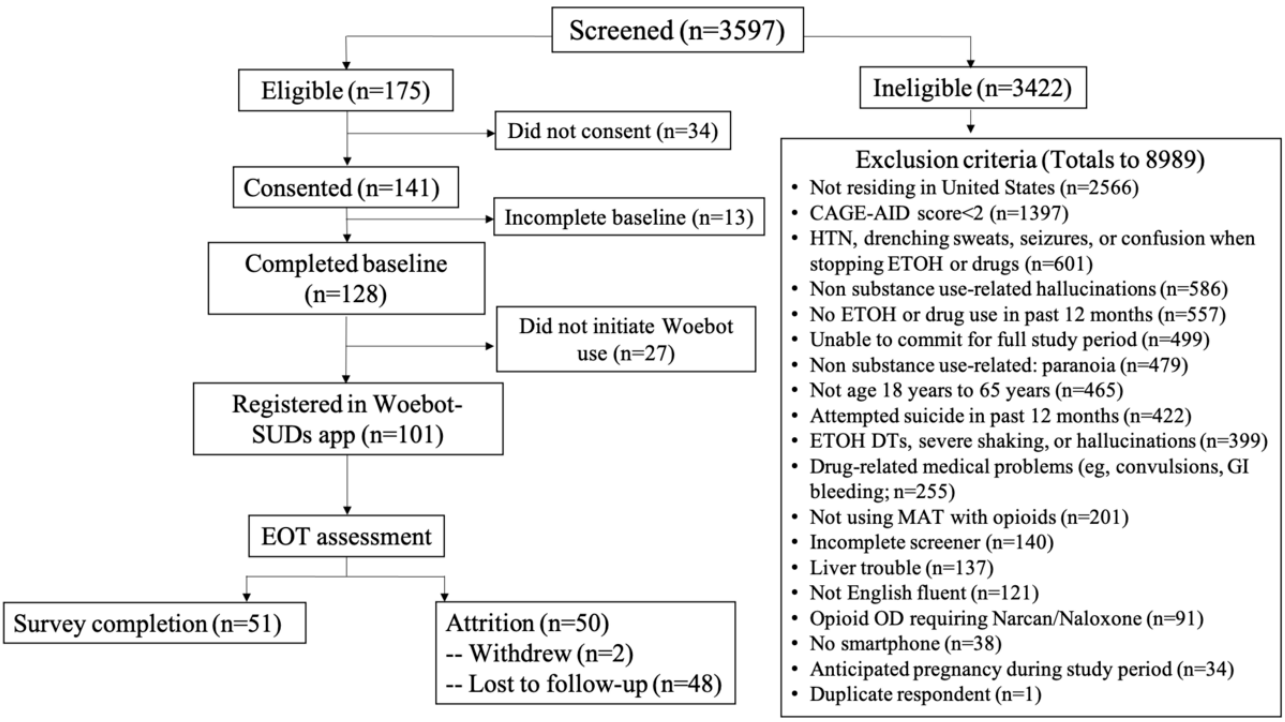
Sample Recruitment

Participants were recruited via the Woebot app, social media (eg, Facebook and Nextdoor), Craigslist, and Stanford staff and student wellness listservs. In addition, study flyers were posted in the San Francisco Bay Area, and email invitations were sent to participants from previous studies. Recruitment materials included the URL on a webpage describing the study for people with substance use concerns. Informed consent was required to screen for eligibility. Those who screened as eligible were asked to provide informed consent for participation in the study.

Inclusion criteria were all genders, aged 18 years to 65 years, residing in the United States, screening positive on the 4-item Cut down, Annoyed, Guilty, Eye opener-Adapted to Include Drugs (CAGE-AID) [30] (ie, score of 2 or higher), owning a smartphone for accessing Woebot, available for the 8-week study, willing to provide an email address, and English literate. The CAGE-AID has demonstrated validity, with high internal consistency in screening for problematic drug and alcohol use; a cutoff point of 2+ on the CAGE-AID has a sensitivity of 70% and specificity of 85% for identifying individuals with SUDs [30]. Study exclusion criteria were current pregnancy, history of severe alcohol or drug-related medical problems (eg, delirium tremens, seizure, liver disease, and hallucinations), opioid overdose requiring Narcan (naloxone), current opioid misuse without medication-assisted treatment, or attempted suicide within the past year.

For this study, the target sample size was 50 participants; however, due to a high level of response and efficiency, enrollment was more than double our recruitment goal. Between March 27, 2020 and May 6, 2020, 3597 individuals were screened for study participation, with 3422 ineligible and 175 eligible individuals. Figure 1 shows the reasons for study exclusion, most frequently residing outside of the United States (2566/3433, 74.75%) and endorsing fewer than 2 criteria on the CAGE-AID (1397/3433, 40.69%). Of the 175 eligible participants, 141 provided informed consent to participate in the study, of whom 128 completed the baseline survey. The analytic sample consisted of 101 participants who ultimately registered with W-SUDs and initiated use. Among the 101 participants enrolled, 11 (10.9%) reported previous use of the Woebot app.

Figure 1. Study consort diagram. CAGE-AID: Cut down, Annoyed, Guilty, Eye Opener-Adapted to Include Drugs; DTs: delirium tremens; EOT: end of treatment; ETOH: ethyl alcohol; HTN: hypertension; MAT: medication-assisted treatment; OD: overdose; Woebot-SUDs: Woebot for the treatment of substance use disorders.



Procedures

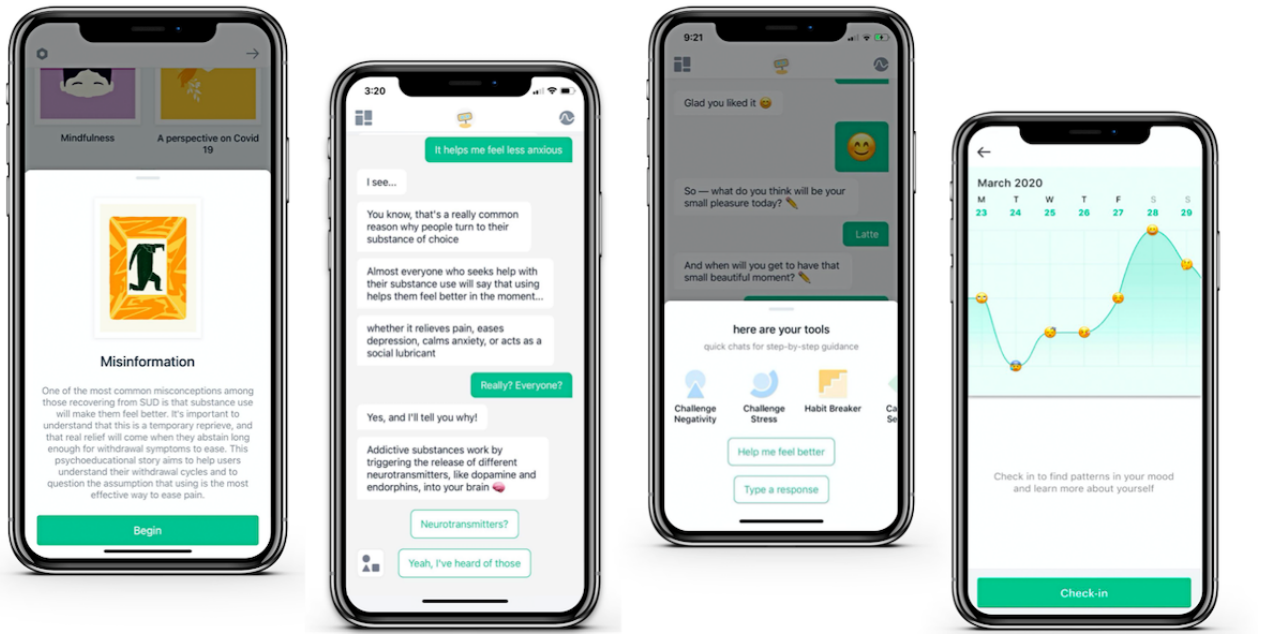
Those who provided informed consent and enrolled were asked to use W-SUDs for 8 weeks. Assessments were administered via Qualtrics at the beginning and end of the 8-week treatment period. Participants received a US $25 Amazon gift card at the end of the study for completing the posttreatment assessment.

W-SUDs Intervention

Described in detail previously [18], Woebot is an automated conversational agent that delivers CBT in the format of brief, daily text-based conversations. The Woebot program is deployed through its own native apps on both iPhone and Android smartphones or devices. The app onboarding process introduces the automated conversational agent, explains the intended use of the device, how data are treated, and the limitations of the service (eg, it is not a crisis service). The user experience is centered around mood tracking and goal-oriented, tailored conversations that can, depending on user input and choice, focus on CBT psychoeducation, application of psychotherapeutic skills for change (eg, thought-challenging), mindfulness exercises, gratitude journaling, and/or reflecting upon patterns and lessons already covered. Each interaction begins with a general inquiry about context (eg, "What's going on in your world right now?") and mood (eg, "How are you feeling?") to ascertain affect in the moment. Additional therapeutic process-oriented features of Woebot include delivery of empathic responses with tailoring to users' stated mood(s), goal setting with regular check-ins for maintaining accountability, a focus on motivation and engagement, and individualized weekly reports to foster reflection. Users become familiar with Woebot, which is a friendly, helpful character that is explicitly not a human or a therapist but rather a guided self-help coach. Daily push notifications prompt users to check in.

We adapted W-SUDs, drawing upon motivational interviewing principles, mindfulness training, dialectical behavior therapy, and CBT for relapse prevention. Sample screenshots from the W-SUDs app are shown in Figure 2. In total, the W-SUDs intervention was developed as an 8-week program with tracking of mood, substance use craving, and pain, with over 50 psychoeducational lessons and psychotherapeutic skills.

Figure 2. Sample screenshots of the Woebot for substance use disorders app: a psychoeducational lesson called Misinformation, the core conversational panel (featuring the Lesson Misinformation), and psychotherapeutic skills for behavior change and mood tracking.



CBT evidence-based, guided self-help treatments have ranged in length from 2 to 12 weeks [31-34], and the National Institutes for Clinical Excellence describes guided self-help as including 6 to 8 face-to-face sessions [35]. Early responsiveness to SUD treatment is predictive of long-term outcomes [36], and brief addiction treatments are efficacious [37]. Brief intervention can minimize potential dropout, a problem common to SUD treatment; [38] therefore, we designed W-SUDs as an 8-week treatment.

Woebot is not designed to address active suicidal ideation or overdose, and this was stated in the study informed consent. In addition, Woebot conversationally informs first-time users that it is not a crisis service. Woebot also has safety net detection that uses natural language processing algorithms to detect and flag several hundred possible harm-to-self phrases (including some misspellings and slang phrases) with 98% accuracy (sensitivity=97 and specificity=99; Woebot Health, unpublished data, September 2020). Woebot detects crisis language (eg, "want to cut myself") and asks to confirm it with the user. If the user confirms, Woebot offers resources (eg, 9-1-1, suicide crisis hotlines), carefully curated with expert consultation. Woebot data indicate that users do not use Woebot for crisis management; approximately 6.3% trigger the safety net protocol, with 27% of those confirming that it is indeed a crisis when Woebot asks to confirm (ie, the true positive rate).

Assessments

Demographic items were assessed at pretreatment; substance use, mental health, and pain measures were administered at pre- and posttreatment; serious adverse events and W-SUDs feasibility and acceptability were assessed at posttreatment; and

W-SUDs use data were collected via the Woebot app over the 8-week intervention. Demographic items included self-reported sex, race and ethnicity, age, marital status, employment status, residential zip code, and sheltering-in-place status given the COVID-19 pandemic.

The Alcohol Use Disorders Identification Test-Concise (AUDIT-C), a widely used 3-item self-report measure based on the 10-item original AUDIT [39], assessed hazardous or harmful alcohol consumption in the past 3 months. A score of 4+ for men and 3+ for women indicated significant problems with alcohol consumption. The AUDIT-C has been found to be a valid screening test for heavy drinking and/or active alcohol abuse or dependence [39]. The Drug Abuse Screening Test-10 (DAST-10), a 10-item self-report measure adapted from the 28-item DAST [40], assessed consequences related to drug abuse, excluding alcohol and tobacco in the past 3 months. The last item of the DAST-10 regarding medical problems resulting from drug use was not reassessed because it was an exclusion criterion in the study screener; hence, the total possible range for the sample was 0-9, not 0-10. Total scores of 3+ indicated significant problems related to drug abuse. The DAST-10 has moderate test-retest reliability, sensitivity, and specificity [40]. For the AUDIT-C and DAST-10 measures at posttreatment, the reference period was the past 2 months, to reflect the period of intervention. Craving was assessed with a single item asking, "In the past 7 days, how much were you bothered by cravings or urges to drink alcohol or use drugs?", with response options of not at all (0), a little bit (1), moderately (2), quite a bit (3), and extremely (4). The Brief Situational Confidence Questionnaire [41], a state-dependent measure, assessed self-confidence to resist the urge "right now" to drink heavily (self-defined) or use drugs in different situations reported on visual analog scales (100 mm lines) anchored from 0% "not at all confident" to 100% "totally confident."

The Patient Health Questionnaire-8 item (PHQ-8), an 8-item scale, assessed depressive symptoms [42], and the Generalized Anxiety Disorder-7 item (GAD-7), a 7-item scale, assessed symptoms of generalized anxiety disorder [43]. Both the PHQ-8 and GAD-7 have good internal consistency and demonstrated convergent validity with measures of depression, stress, and anxiety. A total of 2 items assessed the history of therapy (ever and current) for mental health or substance use concerns. Lifetime psychiatric diagnoses were assessed using 10 items plus a write-in option for others. A single item assessed currently taking prescribed medications for a psychiatric diagnosis.

The treatment feasibility and acceptability of W-SUDs were assessed posttreatment using the Usage Rating Profile-Intervention (URP-I) Feasibility (6 items) and Acceptability (6 items) scales [44], the 8-item Client Satisfaction Questionnaire-8 questions (CSQ-8) [45], and the 12-item Working Alliance Inventory-Short Revised (WAI-SR) [46]. The URP-I item response options ranged from strongly disagree to strongly agree; the items were summed for a total score within each scale, with one feasibility item reverse coded. The CSQ-8 items have 4-point rating scales with response descriptors that vary. Internal consistency exceeds 0.90, and the total sum score ranges from 8 to 32, with higher total scores indicating higher satisfaction. The WAI-SR has three 4-item subscales, with 5-point rating scales, that reflect development of an affective bond in treatment and level of agreement with treatment goals and treatment tasks. Serious adverse events occurring in the 8 weeks after the start of the study were assessed for hospitalization related to substance use, suicide attempt, alcohol or drug overdose, and severe withdrawal (eg, delirium tremens). Positive endorsements were followed up with questions about the timing, diagnosis, and resolution. If additional details were needed to determine whether the event was study related, a team member reached out to the participant. Serious adverse events were reported to the study's Data Safety Monitoring Board (DSMB) within 72 hours of the team learning of the event.

Participants' W-SUDs app use, including days of app use, number of check-ins, and number of messages sent, was collected via the Woebot app, as were module completion rates, lesson acceptability ratings indicated on a binary scale (ie, a thumbs up or thumbs down emoticon), and mood impact after tools utilization (ie, feeling same, better, or worse after completion). In addition, on a daily basis, the W-SUDs app assessed mood, cravings or urges to use, and pain. In-the-moment emotional state was reported through emoji selection with a default menu of 19 total moods, including options for negative (angry, sad, and anxious), positive (happy and content), and average mood (okay), with an additional ability to type in free text emotion words and/or self-selected emoji expressions. Cravings were assessed as not at all (0), a little bit (1), moderately (2), quite a bit (3), or extremely (4). Physical pain was rated on a scale of 0 to 10.

Data Analyses

Descriptive statistics (means and frequencies) were used to describe the sample and examine the ratings of program feasibility and acceptability. Paired samples *t* tests and McNemar nonparametric tests examined within-subject changes from pre- to posttreatment on measures of substance use, confidence, cravings, mood, and pain. Change scores were calculated (pre- minus posttreatment), and bivariate correlations were used to examine associations between changes in AUDIT-C and DAST-10 scores and changes in use occasions, confidence, and depression and anxiety scores. *t* tests were conducted to examine changes from pre- to posttreatment in substance use, confidence, mood, and pain by whether participants were currently in therapy or taking psychiatric medications. Posttreatment survey completion was 50.5% (51/101), with better retention among those with a higher CAGE-AID score at screening (γ=0.37; *P*=.02). Retention was lowest among those with a CAGE-AID score of 2 (7/26, 27%) and higher for those scoring 3 (22/38, 58%) or 4 (22/37, 59%). Retention was unrelated to participant demographic characteristics, previous use of Woebot, psychiatric diagnoses, primary problematic substance, depressive symptoms, pain, cravings, confidence, substance use occasions, AUDIT-C scores, or DAST-10 scores (all *P* values>.102). Missing data on individual survey items was minimal. In a single instance, a participant's average score values were imputed when missing 1 item on the PHQ-8. Participants were prompted to report craving and pain ratings within the W-SUDs app on a daily basis. The data were aggregated so that if participants provided multiple ratings within a day, the scores were averaged. To examine changes over time, generalized estimating equation

linear models were run with week entered as a factor, setting week 1 as the reference category.

Results

Sample Characteristics

Table 1 presents the baseline characteristics of the participants. According to zip code, the sample was drawn from 31 US states, and at baseline, nearly all participants (99/101, 98.0%) reported sheltering in place during the COVID-19 pandemic. Most (73/101, 72.3%) reported a lifetime psychiatric diagnosis, most commonly generalized anxiety disorder (49/101, 48.5%) and unipolar depression (45/101, 44.6%), with 47.5% (48/101) reporting multiple lifetime psychiatric diagnoses; few (6/101, 5.9%) reported a SUD diagnosis, 43.6% (44/101) were currently taking psychiatric medication, and 25.7% (26/101) were currently in therapy.

Table 1. Sample characteristics at baseline (N=101).

Variable	Mean (SD); range	Value, n (%)
Age (years)	36.8 (10.0); 19-62	N/A[a]
Sex		
Female	N/A	76 (75.2)
Male	N/A	25 (24.8)
Race and ethnicity		
Non-Hispanic White	N/A	79 (78.2)
Hispanic/Latinx	N/A	4 (4.0)
Non-Hispanic Black/African-American	N/A	4 (4.0)
Non-Hispanic Asian-American	N/A	3 (3.0)
Multiethnic	N/A	7 (6.9)
Other or missing	N/A	4 (4.0)
Marital status		
Married or cohabitating or partnered	N/A	54 (53.5)
Divorced or separated or widowed	N/A	14 (13.9)
Single or never married	N/A	33 (32.7)
Employment status		
Employed full-time	N/A	62 (61.4)
Employed part-time	N/A	11 (10.9)
Unemployed, job-seeking	N/A	12 (11.9)
Other (eg, retired, disabled, homemaker, and student)	N/A	16 (15.8)
COVID-19 situation		
Sheltering in place, lockdown, quarantined	N/A	99 (98.0)
No restrictions	N/A	2 (2.0)
Lifetime psychiatric diagnoses		
Unipolar depression	N/A	45 (44.6)
Bipolar or manic depression	N/A	10 (9.9)
Anxiety disorder	N/A	49 (48.5)
Posttraumatic stress disorder	N/A	19 (18.8)
Attention deficit hyperactivity disorder	N/A	15 (14.9)
Other (eg, obsessive compulsive disorder, eating disorder, and personality disorder)	N/A	12 (11.9)
Substance use disorder	N/A	6 (5.9)
Multiple psychiatric diagnoses	N/A	48 (47.5)
No lifetime psychiatric diagnoses	N/A	28 (27.7)
Therapy experience		
Never	N/A	30 (29.7)
Formerly	N/A	45 (44.6)
Currently	N/A	26 (25.7)
Currently taking psychiatric medication	N/A	44 (43.6)
Patient Health Questionnaire-8 item depression (possible range 0-24)	10.8 (5.8); 0-24	N/A
10+ moderate-to-severe	N/A	54 (53.5)
General Anxiety Disorder-7 item anxiety (possible range 0-21)	9.6 (5.7); 0-21	N/A
10+ moderate-to-severe	N/A	47 (46.5)

Variable	Mean (SD); range	Value, n (%)
Pain intensity in the past 7 days (possible range 0-100)	20.4 (22.3); 0-80	N/A
Pain interfere with normal work in the past 30 days		
Not at all	N/A	60 (59.4)
A little bit	N/A	23 (22.8)
Moderately	N/A	9 (8.9)
Quite a bit	N/A	7 (6.9)
Extremely	N/A	2 (2.0)
Primary substance		
Alcohol	N/A	69 (68.3)
Cannabis	N/A	20 (19.8)
Stimulants or cocaine	N/A	7 (6.9)
Other (eg, club drugs, pain killers, and sedatives)	N/A	5 (5.0)
Indicated multi-substances	N/A	37 (36.6)
Past 30 days of substance use[b] (days), mean (SD) and n (%) reporting any past 30-day use		
Alcohol	19.4 (9.2); 1-30	88 (87.1)
Cannabis	19.4 (12.2); 1-30	50 (49.5)
Sedatives	5.3 (5.3); 1-15	19 (18.8)
Hallucinogens	2.0 (1.2); 1-5	10 (9.9)
Prescription stimulants	21.9 (11.0); 2-30	10 (9.9)
Cocaine	3.6 (4.0); 1-10	5 (5.0)
Methamphetamine	22.5 (13.1); 3-30	4 (4.0)
Inhalants	5.8 (3.3); 2-10	4 (4.0)
Prescription opioids	12.4 (15.2); 1-30	5 (5.0)
Street opioids	0 (0); 0	0 (0)
Number of substance use occasions in the past 30 days	31.8 (17.7); 0-76	N/A
Alcohol Use Disorders Identification Test-Concise (possible range 0-12)	5.5 (3.1); 0-12	N/A
Men (% with score 4+, clinical range)	5.2 (3.2); 0-11	18 (72)
Women (% with score 3+, clinical range)	5.5 (3.1); 0-12	59 (78)
Drug Abuse Screening Test-10 item (possible range 0-10)	3.0 (2.6); 0-8	N/A
% with score 3+, clinical range	N/A	56 (55.4)
Bothered by cravings in the past 7 days		
Not at all	N/A	7 (6.9)
A little bit	N/A	31 (30.7)
Moderately	N/A	33 (32.7)
Quite a bit	N/A	23 (22.8)
Extremely	N/A	7 (6.9)
Current confidence scores[c] (possible range 0%-100%)		
Negative emotional	38.5 (30.1); 0-100	N/A
Negative physical	50.9 (33.4); 0-100	N/A
Positive emotional	60.5 (31.7); 0-100	N/A
Testing personal control	54.1 (34.2); 0-100	N/A
Urges and temptations	41.1 (28.7); 0-100	N/A

Variable	Mean (SD); range	Value, n (%)
Interpersonal conflict	44.4 (30.9); 0-100	N/A
Social pressure	49.2 (33.4); 0-100	N/A
Positive social	46.0 (31.2); 0-100	N/A
Overall confidence average score	48.1 (22.1); 0-100	N/A

[a]N/A: not applicable.

[b]Six participants reported no substance use in the past 30 days at baseline. The mean days of use were calculated among those who reported any use of that substance in the past 30 days.

[c]Values presented are percentages.

Substance Use at Pretreatment

Self-identified primary problematic substances were alcohol (69/101, 68.3%), cannabis (20/101, 19.8%), stimulants or cocaine (7/101, 6.9%), and other (5/101, 4.9%). Over a third (37/101, 36.6%) indicated problems with multiple substances. Most (88/101, 87.1%) reported use of alcohol in the past month; among past month drinkers, alcohol use averaged 19.4 of the past 30 days. About half (50/101, 49.5%) reported use of cannabis in the past month, among users averaging 19.4 of the past 30 days. Less common was use of sedatives (19/101, 18.8%), hallucinogens (10/101, 9.9%), and prescription stimulants (10/101, 9.9%) in the past month. None of the participants reported use of street opioids in the past month. Combining reported days of use across substances, the number of use occasions in the past 30 days averaged 31.8 (SD 17.7) with a wide range of 0-76. At baseline, AUDIT-C scores averaged 5.5 (SD 3.1) for the overall sample, with 72% (18/25) of men and 78% (59/76) of women scoring in the clinical range. DAST-10 scores averaged 3.0 (SD 2.6), with 55.4% (56/101) scoring in the clinical range. Nearly two-thirds (63/101, 62.4%) of the sample reported being bothered in the past 7 days by moderate-to-extreme cravings or urges to drink alcohol or use drugs. Participants' confidence in 8 domains to resist urges to use substances ranged from an average of 60.5% (SD 31.7) for positive emotional states to 38.5% (SD 30.1) for negative emotional states, with an overall average of 48.1% (SD 22.1) and a wide range of 1%-100%.

W-SUDs Use and Within-App User Feedback

Among the full sample (N=101), for the 8-week treatment period, participants' use of W-SUDs averaged 15.7 days (SD 14.2; median 10; IQR 20) or 2.0 times per week, with an average of 600.7 user sent messages (SD 556.5; median 360; IQR 763) or 75.1 messages per week and engagement on average with 12.1 modules (SD 8.3; median 9; IQR 12.5), which consist of psychoeducational lessons and psychotherapeutic tools for mood and behavior change. An indicator of intervention engagement over time, Multimedia Appendix 1 shows the percentage of participants actively sending messages by treatment week and, among those participating each week, their average number of messages. The types of conversations vary in length; therefore, the total number of messages sent does not necessarily reflect the richness of content reviewed. In addition, the individuals in each week are not necessarily the same across weeks. For example, someone could have sent messages in weeks 2 to 4 and 6 to 7 but not in weeks 5 or 8. The sample completed an average of 7.9 psychoeducational lessons (SD 7.6; median 4; IQR 12). Lesson completion rates were highest (>50%) for content concerning COVID-19, urge surfing, and SUD labels and lowest (<5%) for content concerning sleep and grief. Lesson acceptability ratings were high across the board, with 94.0% (562/598) of completed lessons receiving thumbs up. Participants used an average of 4.3 tools (SD 1.4; median 4; IQR 1). Mood impact after tool utilization, denoting in-vivo mood modulation, was predominately positive (better=70%, same=24%, and worse=6%). In total, 14 of the 101 users (13.9%) completed all of the psychoeducational lessons in W-SUDs before the end of the 8-week intervention period.

W-SUDs Mood, Craving, and Pain Ratings

A total of 1571 mood ratings were entered into the W-SUDs app by 90 of the 101 (89.1%) participants, with each participant entering on average 17.5 mood ratings (SD 16.1; median 10; IQR 25.3) or 2.2 per week. A total of 1399 craving and 1403 pain ratings were entered into the W-SUDs app by 87 of the 101 participants (86.1%), with each participant providing an average of 16.1 ratings (SD 14.8; median 9; IQR 21) for cravings and 16.1 ratings (SD 14.9; median 9; IQR 21) for pain. Table 2 shows the number of participants providing craving ratings for each week and summarizes the generalized estimating equation model analyzing craving ratings over time. Compared with week 1, craving ratings were significantly lower at weeks 4 through 9. By weeks 8 and 9, craving ratings were reduced by approximately half of the sample's mean rating at week 1. In contrast, pain ratings did not differ significantly by week and over the 9 weeks averaged 2.3 (SD 2.1), on a scale of 0 to 10.

Table 2. Participants' (N=101) craving ratings from week 1 to week 9 reported in the Woebot for the treatment of substance use disorders (W-SUDs) app.

Variable	Value, n[a] (%)	Craving, mean (SD)[b]	β	SE	Wald X^2 (*df*)	*P* value	Exp (B)	95% CI
Week 1[c]	82 (81.2)	1.59 (0.11)	0[d]	Ref	Ref	Ref	Ref	Ref
Week 2	69 (68.3)	1.48 (0.12)	−.11	0.09	1.35 (1)	.25	0.90	0.761.08
Week 3	55 (54.5)	1.32 (0.14)	−.27	0.14	3.59 (1)	.06	0.77	0.58-1.01
Week 4	48 (47.5)	1.21 (0.17)	−.38	0.16	5.33 (1)	.02	0.69	0.50-0.95
Week 5	39 (38.6)	0.88 (0.15)	−.71	0.16	21.21 (1)	<.001	0.49	0.36-0.66
Week 6	32 (31.7)	1.01 (0.21)	−.58	0.20	8.46 (1)	.004	0.56	0.38-0.83
Week 7	30 (29.7)	0.98 (0.18)	−.61	0.20	9.19 (1)	.002	0.54	0.37-0.81
Week 8	24 (23.8)	0.81 (0.19)	−.78	0.21	13.84 (1)	<.001	0.46	0.30-0.69
Week 9[e]	20 (19.8)	0.86 (0.2)	−.73	0.21	11.62 (1)	.001	0.48	0.32-0.73

[a]Number of participants reporting their craving at least once each week with response options of not at all (0), a little bit (1), moderately (2), quite a bit (3), or extremely (4).

[b]Model estimated marginal means (SD).

[c]Week 1 is the reference group to which all other weeks are compared.

[d]Set to zero as the reference category.

[e]Woebot for the treatment of substance use disorders is offered as an 8-week treatment; however, participants could continue to use the app.

Changes Pre- to Posttreatment

Table 3 shows scores for the participants who completed assessments at both pre- and posttreatment. In paired sample *t* tests, confidence scores overall and in all 8 domains significantly increased from pre- to posttreatment (all *P* values<.05). In addition, significant reductions were observed from pre- to posttreatment in past month substance use occasions, AUDIT-C and DAST-10 scores (overall and among those in the clinical range at pretreatment), and PHQ-8 depression and GAD-7 anxiety scores (all *P* values<.05). A McNemar test indicated significant reductions in cravings, with more participants reporting little to no cravings and fewer reporting moderate-to-extreme cravings from pre- to posttreatment (*P*<.001). Reports of pain intensity and pain interference with work did not change significantly from pre- to posttreatment.

A greater decline in the AUDIT-C score was associated with greater reductions in use occasions (*r*=0.48), PHQ-8 depression (*r*=0.36), and GAD-7 anxiety (*r*=0.34) scores and with increases in confidence (*r*=−0.39; all *P* values<.02). A greater decline in the DAST-10 score was associated with greater reductions in PHQ-8 depression (*r*=0.40; *P*<.01) but not with the number of use occasions (*r*=0.10), confidence (*r*=−0.12), or GAD-7 anxiety (*r*=0.21).

Of the 14 *t* tests, only 1 was statistically significant as to whether participants currently in therapy or taking psychiatric medications showed greater pre- to posttreatment changes in substance use (use occasions, AUDIT-C, and DAST-10), confidence, mood (PHQ-8 and GAD-7), or pain. The finding was that participants currently in therapy reported greater reductions from pre- to posttreatment in depressive symptoms (n=16; mean change −4.7, SD 4.5) than those not currently in therapy (n=35; mean change −0.9, SD 5.1; t_{49}=2.55; *P*=.01).

Table 3. Pre- to posttreatment changes in substance use and mental health measures (n=51).

Variable	Pretreatment	Posttreatment	*t* test (*df*)	*P* value
Substance use occasions[a], mean (SD)	29.5 (14.0)	20.1 (17.8)	−4.72 (50)	<.001
Alcohol Use Disorders Identification Test-Concise, mean (SD)				
Full sample	5.3 (2.9)	4.0 (3.2)	−3.58 (50)	<.001
At-risk at pretreatment[b]	6.7 (2.0)	4.9 (3.2)	−3.92 (38)	<.001
Drug Abuse Screening Test-10 item, mean (SD)				
Full sample	2.9 (2.7)	1.7 (2.4)	−4.25 (50)	<.001
At-risk at pretreatment[b]	5.3 (1.5)	3.1 (2.6)	−5.00 (26)	<.001
Confidence scores (0%-100%), mean (SD)				
Negative emotional	37.2 (28.4)	56.3 (27.9)	3.86 (50)	<.001
Negative physical	49.2 (31.1)	64.8 (30.7)	3.41 (50)	.001
Positive emotional	61.7 (28.2)	75.2 (26.8)	2.98 (50)	.004
Testing personal control	51.7 (31.0)	63.4 (30.6)	2.42 (50)	.03
Urges and temptations	37.9 (23.6)	57.5 (28.6)	5.52 (50)	<.001
Interpersonal conflict	40.9 (28.2)	61.2 (31.1)	4.62 (50)	<.001
Social pressure	43.8 (32.5)	63.4 (33.7)	3.64 (50)	.001
Positive social	45.5 (31.6)	60.8 (30.2)	3.56 (50)	.001
Overall confidence average score	46.0 (19.3)	62.8 (22.4)	5.62 (50)	<.001
Bothered by cravings in the past 7 days, n (%)				
Not at all or a little bit	16 (31.4)	27 (53)	N/A[c]	.013[d]
Moderately or quite a bit or extremely	35 (68.6)	24 (47)	N/A	.013[d]
Pain intensity in the past 7 days, mean (SD)	24.3 (22.2)	24.2 (22.0)	−0.02 (50)	.982
Pain interfere with work in the past 30 days, n (%)				
Not at all or a little bit	40 (78.4)	40 (78)	N/A	1.00[d]
Moderately or quite a bit or extremely	11 (21.5)	11 (22)	N/A	1.00[d]
Patient Health Questionnaire-8 item depression, mean (SD)	10.7 (5.3)	8.6 (5.1)	−2.91 (50)	.005
General Anxiety Disorder-7 item anxiety, mean (SD)	10.1 (5.7)	7.8 (5.3)	−3.45 (50)	.001

[a]Reflects number of days of use summed across substances.

[b]Analyses run for the subgroup of participants scoring in the clinical range at pretreatment, which are scores of 4+ for men and 3+ for women on the Alcohol Use Disorders Identification Test-Concise and scores of 3+ on the Drug Abuse Screening Test-10 item.

[c]N/A: not applicable.

[d]*P* value obtained with McNemar's test.

Serious Adverse Events

Among the 51 participants who completed the posttreatment assessment, 1 reported a serious adverse event. An individual reported hospitalization for treatment of sepsis secondary to switching from smoking to injecting illicit drugs, shortly before or at the start of study participation, and was deemed by the DSMB to be unrelated to study involvement.

Feasibility and Acceptability Ratings

Table 4 shows the mean scores and ranges of the 4 feasibility and acceptability measures completed posttreatment. On the individual CSQ-8 items, the majority (35/51, 69%) indicated that they would return to the program, reported that interactions with W-SUDs helped them deal more effectively with their problems (35/51, 69%), were mostly or very satisfied overall (36/51, 71%), were satisfied with the amount of help received (37/51, 73%), rated the quality of interaction on W-SUDs as good or excellent (39/51, 76%), would recommend W-SUDs to a friend (39/51, 76%), and received the kind of service they wanted (41/51, 80%). A lower percentage of participants stated that W-SUDs met most or all of their needs (22/51, 43%). Scores for the 3 WAI-SR subscales, with identical response options, differed significantly from each other in pairwise *t* test comparisons (all *P* values<.05), with the highest ratings on development of an affective bond to Woebot, followed by agreement on the tasks of treatment and then agreement on the goals of treatment.

Table 4. Woebot for the treatment of substance use disorders (W-SUDs) posttreatment feasibility and acceptability ratings (n=51).

Measure	Mean (SD); range
Usage Rating Profile-Intervention (12 items, possible range 12-72)	
Feasibility (6 items, range 6-36)	28.5 (5.7); 11-36
Acceptability (6 items, range 6-36)	25.6 (7.3); 6-36
Client Satisfaction Questionnaire (8 items, possible range 8-32)	23.2 (5.5); 8-31
Working Alliance Inventory-Short Revised (12 items, possible range 12-60)	40.8 (12.5); 12-60
Goal agreement (4 items, range 4-20)	12.4 (4.4); 4-20
Task agreement (4 items, range 4-20)	13.0 (4.8); 4-20
Affective bond formation (4 items, range 4-20)	15.4 (4.2); 4-20

CSQ-8 satisfaction scores did not differ by any measured participant characteristics, including sex, race or ethnicity, marital and employment status, age, primary substance of abuse, or history of a psychiatric diagnosis. CSQ-8 satisfaction scores also did not differ by baseline measures of depression, anxiety, pain, craving, confidence, substance use occasions, AUDIT-C, or DAST-10 scores. Non-Hispanic White participants had higher URP-I-Acceptability ratings ($F_{1,50}$=8.32; P=.006) and higher WAI-SR scores ($F_{1,50}$=5.08; P=.03) than participants from other racial or ethnic groups. In addition, URP-I-Acceptability ratings were higher among participants who reported moderate-to-extreme craving at baseline ($F_{1,50}$=5.21; P=.03). Finally, older age (r=0.36; P=.01) and reporting of moderate-to-extreme impairment due to pain at baseline ($F_{1,50}$=4.36; P=.04) were associated with higher URP-I-Feasibility ratings.

A greater reduction in substance use occasions from pre- to posttreatment was significantly associated with higher WAI-SR (r=−0.37; P=.008) and URP-I-Acceptability (r=−0.30; P=.03) scores. An increase in confidence to resist urges to use substances was also associated with higher scores on the WAI-SR (r=0.30; P=.03), URP-I-Acceptability (r=0.33; P=.02), and CSQ-8 (r=0.28; P=.045). Changes in AUDIT-C, DAST-10, depression, and anxiety measures were not associated with acceptability and feasibility ratings.

Discussion

Principal Findings

W-SUDs, an automated conversational agent, was feasible to deliver, engaging, and acceptable and was associated with significant improvements pre- to posttreatment in self-reported measures of substance use, confidence, craving, depression, and anxiety and in-app measures of craving. The W-SUDs app registration rate among those who completed the baseline survey was 78.9% (101/128), comparable with other successful mobile health interventions [47]. As expected, the use of the W-SUDs app was highest early in treatment and declined over the 8 weeks. Study of engagement with digital health apps has been growing, with no consensus yet on ideal construct definitions [48-50]. Simply reporting the number of messages or minutes spent on an app over time may undermine clarity and genuine understanding of the type and manifestation of app utilization related to clinical outcomes of interest [51]. Further research in this area is warranted.

The observed reductions from pre- to posttreatment measures of depression and anxiety symptoms were consistent with a previous evaluation of Woebot conducted with college students self-identified as having symptoms of anxiety and depression [18]. Furthermore, in this study, treatment-related reductions in depression and anxiety symptoms were associated with declines in problematic substance use. Declines in depressive symptoms observed from pre- to posttreatment were greater among the participants in therapy.

This study also examined working alliance, proposed to mediate clinical outcomes in traditional therapeutic settings [52]. Traditionally, working alliance has been characterized as the cooperation and collaboration in the therapeutic relationship between the patient and the therapist [53-55]. The role of working alliance in relationally based systems and digital therapeutics has been previously considered [16,17,56]; the potential of alliance to mediate outcomes in Woebot should be further validated in future studies adequately powered to examine mediators of change.

Measures of physical pain did not change with the use of W-SUDs as reported in pre- and posttreatment measures or within the app; however, the sample's baseline ratings of pain intensity and pain interference were low. Although not a direct intervention target, pain was measured due to the potential for use of substances to self-treat physical pain and the possibility that pain may worsen if substance use was reduced, which was not observed here.

Within-app lesson completion and content acceptability were high for the overall sample, although there was a wide range of use patterns. Most participants used all facets of the W-SUDs app: tracked their mood, cravings, and pain; completed on average over 7 psychoeducational lessons; and used tools in the W-SUDs app. Only about half of the sample completed the posttreatment assessment, with better retention among those screening higher on the CAGE-AID. That is, those with more severe substance use problems at the start of the study, and hence in greater need of the intervention, were more likely to complete the posttreatment evaluation. None of the other measured variables distinguished those who did and did not complete the posttreatment evaluation. This level of attrition is

commensurate with other digital mental health solution trial attrition rates [47,57].

Comparison With Previous Work

By addressing problematic substance use, including but not limited to alcohol, the W-SUDs intervention supports and extends a growing body of literature on the use of automated conversational agents (or chatbots) and other mobile apps to support behavioral health. A systematic review of mobile and web-based interventions targeting the reduction of problematic substance use found that most web-based interventions produced significant short-term improvements in at least one measure of problematic substance use [6]. Mobile apps were less common than web-based interventions, with weaker evidence of efficacy and some indication of causing harm (ie, inadvertently helping users increase, rather than decrease, their blood alcohol level while partying). However, mobile interventions can be efficacious. Electronic screening and brief intervention programs, which use mobile tools to screen for excessive alcohol use and deliver personalized feedback, have been found to effectively reduce alcohol consumption and alcohol-related problems [58]. However, rigorous evaluation trials of digital interventions targeting nonalcohol substance use are limited [7]. Furthermore, although a systematic review concluded that conversational agents showed preliminary efficacy in reducing psychological distress among adults with mental health concerns compared with inactive control conditions [27], this is the first published study of a conversational agent adapted for substance use.

Study Strengths

Study strengths include study enrollment being double the initial recruitment goal, reflecting interest in W-SUDs. Most participants reported lifetime psychiatric diagnoses, and approximately half of the participants endorsed current moderate-to-severe levels of depression or anxiety. W-SUDs was used on average twice per week during the 8-week program. From pre- to posttreatment with W-SUDs, participants reported significant improvements in multiple measures of substance use and mood. The delivery modality of W-SUDs offered easy, immediate, and stigma-free access to emotional support and substance use recovery information, particularly relevant during a time of global physical distancing and sheltering in place. More time spent at home, coupled with reduced access to in-person mental health care, may have increased enrollment and engagement with the app. Although further data on recruitment and enrollment are warranted, these early findings suggest that individuals with SUDs are indeed interested in obtaining support for this condition from a fully digitalized conversational agent.

Limitations and Future Directions

This study had a single-group design, and the outcomes were short term and limited to posttreatment, thus limiting the strength of inferences that can be drawn. The sample was predominately female and identified as non-Hispanic White, and the majority were employed full-time. Non-Hispanic White participants reported higher program acceptability on 2 of the 4 measures compared with participants from other racial or ethnic groups. Future research on W-SUDs will use a randomized design, with longer follow-up, and focus on recruitment of a more diverse population to better inform racial or ethnic cultural programmatic tailoring, using quotas to ensure racial or ethnic diversity in sampling. Notably, although recruited from across the United States, nearly all participants (99/101, 98.0%) were sheltering in place at the time of study enrollment due to the COVID-19 pandemic, which may have affected substance use patterns and mood as well as interest in a digital health intervention. Notably, however, alcohol sales in the United States increased during the COVID-19 pandemic [59]. The primary outcomes of substance use, cravings, confidence, mood, and program acceptability were standard measures with demonstrated validity and reliability. The limitations were that all were self-reported, and acceptability measures were not open-ended or qualitative. Few participants were misusing opioids, likely due to study exclusion designed to mitigate risk, namely, the requirement of engagement with medication-assisted treatment and no history of opioid overdose requiring Narcan (naloxone). Notably, nearly 1400 people with interest in a program for those with substance use concerns were excluded due to low severity on the CAGE-AID screener. Worth testing is the utility of digital health programs for early intervention on substance misuse that is subsyndromal.

Building upon the findings of this study, future research will evaluate W-SUDs in a randomized controlled trial with a more racially or ethnically diverse sample, balanced on sex and primary problematic substance of use; will employ greater strategies for study retention (eg, increased incentives, obtaining phone contact details, and sending more outreach reminders); and will be conducted during a period with less restrictions on social contacts and physical mobility. Randomized controlled evaluations of conversational agent interventions relative to other treatment modalities are required [27,60].

Conclusions

This study is the first empirical evaluation of an SUD-focused digital therapeutic delivered via a fully automated conversational agent. The therapeutic approach is acceptable, feasible, and safe. The study observed significant reductions in substance use and cravings in the context of population-level shifts in the pattern of substance use during a global pandemic. The scalability and accessibility of an automated program coupled with the growing problem of substance use suggest the potential for an engaging and effective therapeutic to reduce the burden of SUDs. Further research is needed to quantify the adoption potential and population impacts of an efficacious digital therapeutic conversational agent for SUD treatment.

Acknowledgments

This research was funded by the National Institute on Drug Abuse (NIDA) with a Small Business Innovations Research Award (Grant # R44D048712). The sponsor was not involved in any of the research or analyses. The study's contents are solely the

responsibility of the authors and do not necessarily represent the official views of NIDA. The authors acknowledge Amer Raheemullah, MD, for providing his addiction medicine expertise to inform the study inclusion and exclusion criteria. The authors acknowledge the study's DSMB members Travis Cook, PhD; Julie Hamlin, LCSW; Danielle Ramo, PhD; and Kelly Young-Wolff, PhD, MPH, for their oversight of the study procedures and outcomes. The authors are grateful for the ongoing guidance from the program officer, Leonardo Angelone, PhD, as the team navigates the different phases of the Small Business Innovation Research award, and the authors appreciate Victor Prikhodko, MBA, for encouraging applications to the Small Business Innovations Research Award mechanism.

Authors' Contributions

JP and AR designed the study, acquired funding, supervised the study recruitment and implementation, and coordinated the research activities. EV and MK provided intellectual input to the study assessments, and MK aided in the development of intervention content. AC led institutional approvals. AC and SP led study recruitment, data accrual, and data management. JP, AC, EV, and MB had access to the study data downloaded from Qualtrics. JP, EV, and MB performed the data analyses. JP drafted the manuscript and incorporated feedback from coauthors.

Conflicts of Interest

AR and SP are employees of Woebot Health. All other authors declare no conflicts of interest related to this study.

Multimedia Appendix 1
Percentage of participants sending messages to Woebot for the treatment of substance use disorders and the average number of messages sent by them each week. W-SUDs: Woebot for the treatment of substance use disorders.
[PNG File , 122 KB - jmir_v23i3e24850_app1.png]

References

1. World Health Organization. Global status report on alcohol and health. Geneva: World Health Organization; 2018.
2. Institute for health metrics and evaluation. Global burden of disease study 2016 results database. 2017. URL: http://ghdx.healthdata.org/gbd-results-tool [accessed 2021-02-17]
3. Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: results from the 2018 National Survey on Drug Use and Health. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. 2019. URL: https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.pdf [accessed 2021-02-17]
4. Grant BF, Stinson FS, Dawson DA, Chou SP, Dufour MC, Compton W, et al. Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: results from the National Epidemiologic Survey on alcohol and related conditions. Arch Gen Psychiatry 2004 Aug;61(8):807-816. [doi: 10.1001/archpsyc.61.8.807] [Medline: 15289279]
5. Keane H. Facing addiction in America: The Surgeon General's Report on Alcohol, Drugs, and Health U.S. Department of Health and Human Services, Office of the Surgeon General Washington, DC, USA: U.S. Department of Health and Human Services, 2016 382 pp. online (grey literature): https://addiction.surgeongeneral.gov/. Drug Alcohol Rev 2018 Feb;37(2):282-283. [doi: 10.1111/dar.12578] [Medline: 29388330]
6. Giroux I, Goulet A, Mercier J, Jacques C, Bouchard S. Online and mobile interventions for problem gambling, alcohol, and drugs: a systematic review. Front Psychol 2017;8:954 [FREE Full text] [doi: 10.3389/fpsyg.2017.00954] [Medline: 28649211]
7. Boumparis N, Schulte MHJ, Riper H. Digital mental health for alcohol and substance use disorders. Curr Treat Options Psych 2019 Nov 26;6(4):352-366. [doi: 10.1007/s40501-019-00190-y]
8. Perrin A, Turner E. Smartphones help blacks, Hispanics bridge some - but not all - digital gaps with whites. Pew Research Center. 2019. URL: https://www.pewresearch.org/fact-tank/2019/08/20/smartphones-help-blacks-hispanics-bridge-some-but-not-all-digital-gaps-with-whites/ [accessed 2021-02-17]
9. Mobile fact sheet. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-02-17]
10. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
11. Smith A. U.S. smartphone use in 2015. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/04/01/us-smartphone-use-in-2015/ [accessed 2021-02-17]
12. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
13. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9) [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]

14. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
15. Lucas GM, Gratch J, King A, Morency L. It's only a computer: Virtual humans increase willingness to disclose. Computers in Human Behavior 2014 Aug;37:94-100 [FREE Full text] [doi: 10.1016/j.chb.2014.04.043]
16. Cook JE, Doyle C. Working alliance in online therapy as compared to face-to-face therapy: preliminary results. Cyberpsychol Behav 2002 Apr;5(2):95-105. [doi: 10.1089/109493102753770480] [Medline: 12025884]
17. Berry K, Salter A, Morris R, James S, Bucci S. Assessing therapeutic alliance in the context of mHealth interventions for mental health problems: development of the mobile agnew relationship measure (mARM) questionnaire. J Med Internet Res 2018 Apr 19;20(4):e90 [FREE Full text] [doi: 10.2196/jmir.8252] [Medline: 29674307]
18. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
19. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
20. Tielman ML, Neerincx MA, Bidarra R, Kybartas B, Brinkman W. A therapy system for post-traumatic stress disorder using a virtual agent and virtual storytelling to reconstruct traumatic memories. J Med Syst 2017 Aug;41(8):125 [FREE Full text] [doi: 10.1007/s10916-017-0771-y] [Medline: 28699083]
21. Gardiner PM, McCue KD, Negash LM, Cheng T, White LF, Yinusa-Nyahkoon L, et al. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: a feasibility randomized control trial. Patient Educ Couns 2017 Sep;100(9):1720-1729 [FREE Full text] [doi: 10.1016/j.pec.2017.04.015] [Medline: 28495391]
22. Bickmore TW, Puskar K, Schlenk EA, Pfeifer LM, Sereika SM. Maintaining reality: relational agents for antipsychotic medication adherence. Interact Comput 2010 Jul;22(4):276-288. [doi: 10.1016/j.intcom.2010.02.001]
23. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a relational agent by hospital patients with depressive symptoms. Interact Comput 2010 Jul 01;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
24. Lucas GM, Rizzo A, Gratch J, Scherer S, Stratou G, Boberg J, et al. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. Front Robot AI 2017 Oct 12;4. [doi: 10.3389/frobt.2017.00051]
25. Philip P, Micoulaud-Franchi J, Sagaspe P, Sevin ED, Olive J, Bioulac S, et al. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. Sci Rep 2017 Feb 16;7 [FREE Full text] [doi: 10.1038/srep42656] [Medline: 28205601]
26. Tielman ML, Neerincx MA, van Meggelen M, Franken I, Brinkman W. How should a virtual agent present psychoeducation? Influence of verbal and textual presentation on adherence. Technol Health Care 2017 Dec 04;25(6):1081-1096 [FREE Full text] [doi: 10.3233/THC-170899] [Medline: 28800346]
27. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. JMIR Ment Health 2019 Oct 18;6(10) [FREE Full text] [doi: 10.2196/14166] [Medline: 31628789]
28. NIAAA council approves definition of binge drinking. NIAAA Newsletter. 2004. URL: https://pubs.niaaa.nih.gov/publications/Newsletter/winter2004/Newsletter_Number3.pdf [accessed 2021-02-17]
29. Miner AS, Shah N, Bullock KD, Arnow BA, Bailenson J, Hancock J. Key considerations for incorporating conversational AI in psychotherapy. Front Psychiatry 2019 Oct 18;10:746 [FREE Full text] [doi: 10.3389/fpsyt.2019.00746] [Medline: 31681047]
30. Brown RL, Rounds LA. Conjoint screening questionnaires for alcohol and other drug abuse: criterion validity in a primary care practice. Wis Med J 1995;94(3):135-140. [Medline: 7778330]
31. Williams C, Wilson P, Morrison J, McMahon A, Walker A, Andrew W, et al. Guided self-help cognitive behavioural therapy for depression in primary care: a randomised controlled trial. PLoS One 2013 Jan 11;8(1) [FREE Full text] [doi: 10.1371/journal.pone.0052735] [Medline: 23326352]
32. Salomonsson S, Santoft F, Lindsäter E, Ejeby K, Ingvar M, Öst LG, et al. Predictors of outcome in guided self-help cognitive behavioural therapy for common mental disorders in primary care. Cogn Behav Ther 2020 Nov 22;49(6):455-474. [doi: 10.1080/16506073.2019.1669701] [Medline: 31638472]
33. Cully JA, Teten AL. A therapist's guide to brief cognitive behavioral therapy. Department of Veterans Affairs, South Central Mental Illness Research, Education, and Clinical Center (MIRECC). 2008. URL: https://depts.washington.edu/dbpeds/therapists_guide_to_brief_cbtmanual.pdf [accessed 2021-02-17]
34. Barth J, Munder T, Gerger H, Nüesch E, Trelle S, Znoj H, et al. Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. PLoS Med 2013;10(5) [FREE Full text] [doi: 10.1371/journal.pmed.1001454] [Medline: 23723742]
35. Appendix E: Glossary. National Institute for Health and Care Excellence; 2011. National Institute for Health and Care Excellence. 2011. URL: https://www.nice.org.uk/guidance/cg123 [accessed 2021-02-17]
36. Miller WR. Are alcoholism treatments effective? The Project MATCH data: response. BMC Public Health 2005 Jul 18;5(1):76 [FREE Full text] [doi: 10.1186/1471-2458-5-76] [Medline: 16026611]

37. DiClemente CC, Corno CM, Graydon MM, Wiprovnick AE, Knoblach DJ. Motivational interviewing, enhancement, and brief interventions over the last decade: a review of reviews of efficacy and effectiveness. Psychol Addict Behav 2017 Dec;31(8):862-887. [doi: 10.1037/adb0000318] [Medline: 29199843]
38. Brorson HH, Arnevik AE, Rand-Hendriksen K, Duckert F. Drop-out from addiction treatment: a systematic review of risk factors. Clin Psychol Rev 2013 Dec;33(8):1010-1024. [doi: 10.1016/j.cpr.2013.07.007] [Medline: 24029221]
39. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
40. Skinner HA. The drug abuse screening test. Addictive Behaviors 1982 Jan;7(4):363-371 [FREE Full text] [doi: 10.1016/0306-4603(82)90005-3]
41. Breslin FC, Sobell LC, Sobell MB, Agrawal S. A comparison of a brief and long version of the situational confidence questionnaire. Behav Res Ther 2000 Dec;38(12):1211-1220. [doi: 10.1016/s0005-7967(99)00152-7] [Medline: 11104185]
42. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173 [FREE Full text] [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
43. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
44. Briesch AM, Chafouleas SM, Neugebauer SR, Riley-Tillman TC. Assessing influences on intervention implementation: revision of the usage rating profile-intervention. J Sch Psychol 2013 Feb;51(1):81-96 [FREE Full text] [doi: 10.1016/j.jsp.2012.08.006] [Medline: 23375174]
45. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Prog Plan 1979 Jan;2(3):197-207 [FREE Full text] [doi: 10.1016/0149-7189(79)90094-6]
46. Hatcher RL, Gillaspy JA. Development and validation of a revised short version of the working alliance inventory. Psychoth Res 2006 Jan;16(1):12-25. [doi: 10.1080/10503300500352500]
47. Cliffe B, Croker A, Denne M, Smith J, Stallard P. Digital cognitive behavioral therapy for insomnia for adolescents with mental health problems: feasibility open trial. JMIR Ment Health 2020 Mar 03;7(3) [FREE Full text] [doi: 10.2196/14842] [Medline: 32134720]
48. O'Brien H. Theoretical perspectives on user engagement. In: Why Engagement Matters. Switzerland: Springer; 2016:1-26.
49. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
50. Holdener M, Gut A, Angerer A. Applicability of the user engagement scale to mobile health: a survey-based quantitative study. JMIR Mhealth Uhealth 2020 Jan 03;8(1) [FREE Full text] [doi: 10.2196/13244] [Medline: 31899454]
51. Chien I, Enrique A, Palacios J, Regan T, Keegan D, Carter D, et al. A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. JAMA Netw Open 2020 Jul 01;3(7) [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10791] [Medline: 32678450]
52. Safran JD, Muran JC. The resolution of ruptures in the therapeutic alliance. J Consult Clin Psychol 1996 Jun;64(3):447-458. [doi: 10.1037//0022-006x.64.3.447] [Medline: 8698936]
53. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy: Theory, Research & Practice 1979;16(3):252-260. [doi: 10.1037/h0085885]
54. Horvath AO, Symonds BD. Relation between working alliance and outcome in psychotherapy: a meta-analysis. J Counsel Psychol 1991;38(2):139-149. [doi: 10.1037/0022-0167.38.2.139]
55. Ruwaard J, Schrieken B, Schrijver M, Broeksteeg J, Dekker J, Vermeulen H, et al. Standardized web-based cognitive behavioural therapy of mild to moderate depression: a randomized controlled trial with a long-term follow-up. Cogn Behav Ther 2009 Dec;38(4):206-221. [doi: 10.1080/16506070802408086] [Medline: 19221919]
56. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30 [FREE Full text] [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
57. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020 Feb 17;3(1):21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
58. Preventing excessive alcohol consumption: electronic Screening and Brief Intervention (e-SBI). Community Preventive Services Task Force. 2012. URL: https://safesupportivelearning.ed.gov/resources/preventing-excessive-alcohol-consumption%C2%A0electronic-screening-and-brief-intervention-e-sbi [accessed 2021-02-17]
59. Rebalancing the 'COVID-19 effect' on alcohol sales. URL: https://nielseniq.com/global/en/insights/2020/rebalancing-the-covid-19-effect-on-alcohol-sales/ [accessed 2021-02-17]

60. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]

Abbreviations

AUDIT-C: Alcohol Use Disorders Identification Test-Concise
CAGE-AID: Cut down, Annoyed, Guilty, Eye opener-Adapted to Include Drugs
CBT: cognitive behavioral therapy
CSQ-8: Client Satisfaction Questionnaire-8 questions
DAST-10: Drug Abuse Screening Test-10
DSMB: Data Safety Monitoring Board
GAD-7: General Anxiety Disorder-7 item
PHQ-8: Patient Health Questionnaire-8 item
SUD: substance use disorder
URP-I: Usage Rating Profile-Intervention
WAI-SR: Working Alliance Inventory-Short Revised
W-SUDs: Woebot for the treatment of substance use disorders

Edited by R Kukafka; submitted 07.10.20; peer-reviewed by B Cliffe, V Stara, E Sezgin; comments to author 14.12.20; revised version received 19.01.21; accepted 31.01.21; published 23.03.21.

Please cite as:
Prochaska JJ, Vogel EA, Chieng A, Kendra M, Baiocchi M, Pajarito S, Robinson A
A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study
J Med Internet Res 2021;23(3):e24850
URL: https://www.jmir.org/2021/3/e24850
doi:10.2196/24850
PMID:33755028

Original Paper

Development and Evaluation of a Mobile App Designed to Increase HIV Testing and Pre-exposure Prophylaxis Use Among Young Men Who Have Sex With Men in the United States: Open Pilot Trial

Katie B Biello[1,2,3,4], MPH, PhD[‡]; Jonathan Hill-Rorie[4], MPH; Pablo K Valente[1], MPH, MD; Donna Futterman[5], MD; Patrick S Sullivan[6], DVM, PhD; Lisa Hightow-Weidman[7], MPH, MD; Kathryn Muessig[8], PhD; Julian Dormitzer[4], MSN, NP; Matthew J Mimiaga[1,2,3,4,9], MA, MPH, SCD; Kenneth H Mayer[4,10,11], MD

[1]Department of Behavioral and Social Sciences, Brown University School of Public Health, Providence, RI, United States
[2]Department of Epidemiology, Brown University School of Public Health, Providence, RI, United States
[3]Center for Health Promotion and Health Equity, Brown University School of Public Health, Providence, RI, United States
[4]The Fenway Institute, Fenway Health, Boston, MA, United States
[5]Adolescent AIDS Program, Children's Hospital at Montefiore Medical Center, The Bronx, NY, United States
[6]Department of Epidemiology, Rollins School of Global Public Health, Emory University, Atlanta, GA, United States
[7]Division of Infectious Diseases, UNC School of Medicine, University of North Carolina, Chapel Hill, NC, United States
[8]Department of Health Behavior, Gillings School of Public Health, University of North Carolina, Chapel Hill, NC, United States
[9]Department of Epidemiology, Fielding School of Public Health, University of California, Los Angelas, CA, United States
[10]Harvard Medical School, Boston, MA, United States
[11]Beth Israel Deaconess Medical Center, Boston, MA, United States
[‡]Written on behalf of iTech Team

Corresponding Author:
Katie B Biello, MPH, PhD
Department of Behavioral and Social Sciences
Brown University School of Public Health
121 S Main Street, 8th Floor
Providence, RI, 02912
United States
Phone: 1 (401)8633082
Email: katie_biello@brown.edu

Abstract

Background: HIV disproportionately affects young men who have sex with men (YMSM) in the United States. Uptake of evidence-based prevention strategies, including routine HIV testing and use of pre-exposure prophylaxis (PrEP), is suboptimal in this population. Novel methods for reaching YMSM are required.

Objective: The aim of this study is to describe the development and evaluate the feasibility and acceptability of the MyChoices app, a mobile app designed to increase HIV testing and PrEP use among YMSM in the United States.

Methods: Informed by the social cognitive theory, the MyChoices app was developed using an iterative process to increase HIV testing and PrEP uptake among YMSM. In 2017, *beta* theater testing was conducted in two US cities to garner feedback (n=4 groups; n=28 YMSM). These findings were used to refine MyChoices, which was then tested for initial acceptability and usability in a technical pilot (N=11 YMSM). Baseline and 2-month postbaseline assessments and exit interviews were completed. Transcripts were coded using a deductive approach, and thematic analysis was used to synthesize data; app acceptability and use data were also reported.

Results: The MyChoices app includes personalized recommendations for HIV testing frequency and PrEP use; information on types of HIV tests and PrEP; ability to search for nearby HIV testing and PrEP care sites; and ability to order free home HIV and sexually transmitted infection test kits, condoms, and lube. In theater testing, YMSM described that MyChoices appears useful and that they would recommend it to peers. Participants liked the *look and feel* of the app and believed that the ability to search

for and be *pinged* when near an HIV testing site would be beneficial. Some suggested that portions of the app felt repetitive and preferred using casual language rather than formal or medicalized terms. Following theater testing, the MyChoices app was refined, and participants in the technical pilot used the app, on average, 8 (SD 5.0; range 2-18) times over 2 months, with an average duration of 28 (SD 38.9) minutes per session. At the 2-month follow-up, the mean System Usability Scale (0-100) score was 71 (ie, above average; SD 11.8). Over 80% (9/11) of the participants reported that MyChoices was useful and 91% (10/11) said that they would recommend it to a friend. In exit interviews, there was a high level of acceptability for the content, interface, and features.

Conclusions: These data show the initial acceptability and user engagement of the MyChoices app. If future studies demonstrate efficacy in increasing HIV testing and PrEP uptake, the app is scalable to reach YMSM across the United States.

Trial Registration: Clinicaltrials.gov NCT03179319; https://clinicaltrials.gov/ct2/show/NCT03179319

International Registered Report Identifier (IRRID): RR2-10.2196/10694

(J Med Internet Res 2021;23(3):e25107) doi:10.2196/25107

KEYWORDS

HIV; men who have sex with men; pre-exposure prophylaxis; pilot study; mobile apps; mobile phone; mHealth

Introduction

Background

HIV incidence remains high in the United States among young men who have sex with men (YMSM). In 2018, more than 20% of new HIV infections in the United States were among young people aged between 13 years and 24 years, with YMSM accounting for 83% of newly diagnosed HIV infections in this age group [1]. New HIV infections also disproportionately impact men who have sex with men (MSM) of color. In 2018, more than half (52%) of the new diagnoses of HIV among MSM aged between 13 years and 24 years were seen among Black individuals and 27% were identified among Latinx individuals [1]. In addition to experiencing a high HIV incidence, a higher proportion of YMSM living with HIV do not know that they are infected in comparison with their adult peers [2,3]. Moreover, individuals aged between 13 years and 24 years are less likely to be linked to HIV care upon diagnosis and present lower levels of viral suppression in comparison with older individuals [2,4,5]. Consequently, YMSM will have delays seeking effective treatment and are more likely to transmit HIV to others [6].

Overwhelming evidence shows that routine HIV testing and expanded use of pre-exposure prophylaxis (PrEP) would drastically reduce the population burden of HIV [7-13]; however, uptake of both interventions is suboptimal among young adults. For example, although the Centers for Disease Control and Prevention recommends that sexually active MSM be tested for HIV at least annually [14], data suggest that nearly half of YMSM reported not being tested for HIV in the past year and one-third reported they had never been tested [15]. In addition, PrEP awareness and uptake are low among younger people [16,17]. For example, only 5% of MSM aged between 18 years and 24 years with PrEP indications reported ever using PrEP, compared with 14% of those aged 25 years and more [18]. Furthermore, only 0.1% of PrEP prescriptions in the United States were given to individuals aged under 18 years [19]. Moreover, young individuals have lower levels of adherence to and retention in PrEP care after initial prescription [20-23], decreasing the impact of PrEP on HIV prevention among this group.

Risk taking, behavioral experimentation, and confronting a host of difficult choices regarding identity formation are all part of the normal developmental trajectory of adolescence and young adulthood [24]. In addition, beliefs about invincibility, sensation seeking, and the still-developing cognitive processes of adolescents may have a role in increased HIV risk-taking behaviors and a lower prioritization of prevention strategies for this age group [25-27]. Developing innovative ways to intervene to increase engagement in HIV prevention behaviors among youth is crucial, particularly interventions that are accessible and responsive to the diverse needs of youth.

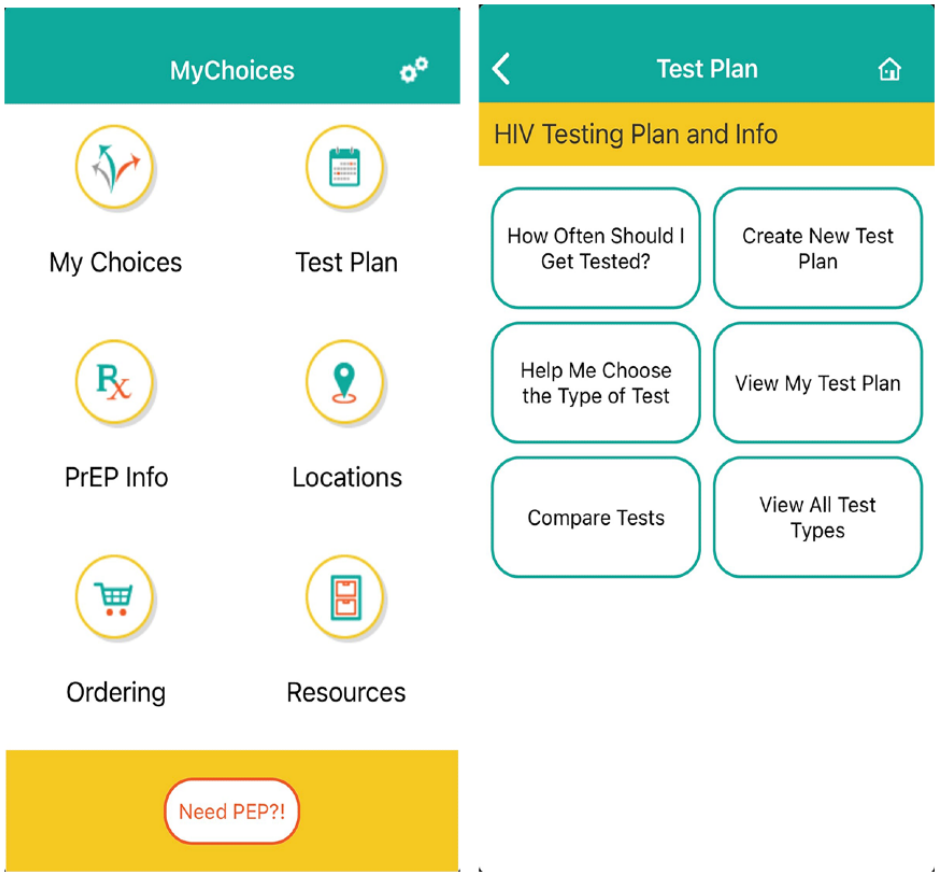
Smartphones are used by nearly all youth in the United States, across race and social class, and as such are a direct way to *meet youth where they are* [28]. In addition, the use of mobile phone apps by YMSM is nearly ubiquitous; apps may offer unique opportunities for public health interventions, and previous studies have demonstrated the feasibility and potential efficacy of this approach [29,30]. In a systematic review of various mobile health (mHealth) interventions, Muessig et al [29] noted that internet- and mobile–based interventions can increase dissemination of HIV prevention interventions to wider populations while also providing consistency and lower cost in intervention delivery once fully developed. In addition, mHealth tools could promote behavior change and improve aspects of the HIV care continuum, including linkage to care, retention in care, and adherence to both PrEP and antiretrovirals [30,31]. As such, a mobile phone app that aims to increase HIV testing and PrEP uptake among YMSM has the potential to provide greater access to and uptake of these prevention services for this population.

MyChoices App

The development of the MyChoices app has been described previously [32]. In brief, MyChoices is a social cognitive theory–driven mobile app adapted from HealthMindr, an HIV prevention app developed through an iterative process for adult populations of MSM [33-35]. MyChoices built upon the initial framework and was subsequently adapted for youth by an interdisciplinary team of researchers with input from a diverse

sample of YMSM at every stage of adaptation and refinement. The goal of the app is to increase HIV testing and PrEP uptake among YMSM in the United States by supporting goal setting, increasing self-efficacy, and enhancing self-regulation [36,37]. Key features of the MyChoices app include facilitation of the development of HIV testing plans with personalized recommendations; inclusion of reminder systems for HIV testing and GPS-enabled maps with local HIV and sexually transmitted infection (STI) testing locations and PrEP providers; ability to order free condoms, condom-compatible lubricants, and at-home HIV and STI test kits; and sexual health information using a variety of media (eg, videos, Graphics Interchange Format [GIF], infographics, frequently asked questions [FAQs], and quizzes) [32]. Figure 1 shows the MyChoices app home screen and the testing plan feature.

Figure 1. MyChoices app screenshots.



Methods

Study Population

Eligible participants were cisgender men who were aged 15-24 years; did not have an HIV test in the past 3 months; self-reported being HIV uninfected or HIV status unknown at screening; owned an iOS or Android mobile phone and were willing and able to download the MyChoices app; were able to understand, read, and speak English; were not taking PrEP; and had self-reported evidence of being at risk for HIV acquisition (details on risk criteria are given in the study by Biello et al [32]).

Through the University of North Carolina/Emory Center for Innovative Technology (iTech) [38], a part of the National Institutes of Health's Adolescent Medicine Trials Network for HIV/AIDS Interventions [39], participants were recruited across 2 sites: Boston, Massachusetts (study site: Fenway Health), and the Bronx, New York City (study site: the Adolescent AIDS Program at Montefiore). Recruitment methods included posting on social media (eg, Craigslist, social networking ads, and gay networking mobile apps); distributing posters, flyers, and palm cards about the study; direct outreach at local venues frequented by YMSM (eg, community-based organizations, schools, bars, health fairs, and balls); and clinic-based recruitment.

Theater Testing for App Refinement

After an initial prototype of the MyChoices app was developed through multiple rounds of formative research with YMSM [33,40], we conducted theater testing with 28 YMSM in 4 groups across the 2 iTech sites (5-8 participants per group). Theater testing allows for groups of participants to interact with the product being tested and provide feedback in situations that approximate real-life experiences and has been used commonly in mHealth app development [33,35,41,42]. Testing was conducted in a private room at each site by a research staff member who had training in qualitative methods and group facilitation. All participants completed a written informed consent or assent process before data collection commenced. Participants completed a brief demographic and behavioral questionnaire before theater testing to contextualize the group data collected. During theater testing, participants interacted with the MyChoices app prototype and provided feedback on

the functionality, appearance, and usability of the platform. We also asked participants to comment on ways to maximize acceptability (eg, update language and improve flow) and to identify the components of the app that were liked or disliked and aspects that could be improved to increase HIV testing and PrEP uptake among YMSM. Groups lasted 60 minutes to 90 minutes and were audio recorded and professionally transcribed verbatim. Participants received US $50 as reimbursement for their time.

Members of the iTech Analytic Core [43] reviewed transcripts for quality and identified emergent themes. We then used Dedoose Version 8.0.35 (SocioCultural Research Consultants, LLC) software to apply the final codes to all transcripts. Thematic analysis involved using a primarily deductive approach to synthesize data coded for app acceptability, particularly around functionality, appearance, usability, and potential for improving HIV testing and PrEP uptake [44,45]. Findings are illustrated in the following sections using representative quotes. These data were used to refine the app before the initiation of the open technical pilot.

Technical Pilot to Assess Feasibility and Acceptability of the MyChoices App

Once the MyChoices app had been refined, a technical pilot with 11 YMSM across the 2 iTech sites was conducted to assess the initial feasibility and acceptability and to identify any final areas for improvement. Eligible participants (mentioned earlier) attended a visit at the study site, at which they completed a web-based behavioral and psychosocial assessment, which included measures to assess sociodemographics (ie, enrollment city, age, race or ethnicity, educational status, and insurance status), sexual behaviors (ie, frequency of condomless anal sex), HIV testing history, and PrEP awareness. Study staff then assisted youth with app download and provided them with brief instructions on the purpose of the MyChoices app and an overview of how to use it; participants were encouraged to use the app over the course of 2 months.

At month 2, the participants completed a web-based assessment. In addition to the measures collected at baseline, we assessed *the acceptability* of the MyChoices app using the System Usability Scale (SUS) [46]. SUS is a validated 10-item measure that assesses the subjective usability of a system or, in this case, an app [46]. SUS has been extensively used in mHealth research and provides reliable results even with small sample sizes [47]. It is scored from 0 to 100, and a score of ≥50 indicates that the app is acceptable [48]. *Feasibility* was assessed using app analytics to determine whether the app was used, how often it was used, and what components were used most and least frequently.

Finally, we conducted *exit interviews* with participants to obtain feedback on app functionality, technical performance, errors and software bugs encountered, overall experiences using the app, feedback for further refinement, and subjective impact of the app on HIV testing and PrEP uptake. Exit interviews were conducted by study staff on the web using videoconferencing technology that was compliant with the Health Insurance Portability and Accountability Act. Interviews were transcribed and analyzed using the same approach outlined above.

Participants received US $50 for completing the baseline visit, US $25 for the 2-month assessment, and US $50 for the exit interview.

The study procedures were reviewed and approved by the University of North Carolina Institutional Review Board (IRB) as a single IRB-of-Record. IRB authorization agreements with all participating research entities were enacted. The MyChoices protocol is registered at ClinicalTrials.gov (NCT03179319).

Results

Findings From Theater Testing

Table 1 provides the characteristics of the 28 YMSM who participated in theater testing. Participants' ages ranged from 16 years to 24 years, with a mean age of 20 years. YMSM of color made up 57% (16/28) of the sample, and 57% (16/28) of the participants were still in school.

A total of four key themes emerged from theater testing, and thus, we structure the presentation of our results to highlight these findings: (1) general utility and acceptability of the app, (2) feedback and suggestions for user interface, (3) opinions on language, and (4) suggestions for additional content and features.

Table 1. Demographic characteristics of participants in theater testing of a novel HIV prevention app for HIV prevention in young men who have sex with men in Boston, Massachusetts, and the Bronx, New York City, 2019 (N=28).

Characteristics	Values
Age (years), mean (SD)	20 (2.0)
Number of condomless anal sex acts in the past 6 months, mean (SD)	10 (15.4)
Study site, n (%)	
Boston, Massachusetts	15 (54)
The Bronx, New York City	13 (46)
Hispanic or Latinx, n (%)	7 (25)
Race, n (%)	
Black	10 (36)
White	12 (43)
Multiracial or other	6 (21)
Currently in school, n (%)	16 (57)
Highest level of education completed, n (%)	
Less than high school	4 (14)
High school diploma or graduate equivalency degree	8 (29)
Some college, or technical or vocational school	13 (46)
Four-year college graduate or more	3 (11)
Currently has health insurance, n (%)	24 (86)
HIV test in the past 3 months, n (%)	19 (68)
Heard of PrEP[a] before the study, n (%)	26 (93)
Ever used internet or apps for the following (not mutually exclusive), n (%)	
Tracking health behaviors	15 (54)
Getting information about HIV or other STDs[b]	13 (46)
Getting other health or medical information	15 (54)
Sending reminders	21 (75)

[a]PrEP: pre-exposure prophylaxis.

[b]STD: sexually transmitted disease.

General Utility and Acceptability of MyChoices

Participants described that an HIV prevention app needs to be perceived as useful to its target audience and that they saw value in MyChoices. The participants described:

> *Yeah, I feel like perceived value...in terms of repeated use, you sort of have to make sure that people get into using it for like reminders or like plans. Otherwise, there's no reason to go back. Otherwise it's just like an information center that you could find on the Internet, you know?* [White, gay, age 21 years, Boston]

> *I think like this app is very, very helpful and useful. Like I like how the questions are not too personal. Like they're the same questions that doctors would ask you. And I like how you can take a survey and it shows when you should get tested, like every three months or something like that. And the ordering, I think that is very smart. And the location, that is very smart too. And I just like – I like this app overall.* [Black, gay, age 16 years, the Bronx]

Although participants acknowledged that they could find similar information from other sources, they appreciated that the information on MyChoices came from a reliable source because of its association with the study and clinic and that using the app was a better avenue for ascertaining health information than searching the internet for resources they could not be sure were accurate:

> *I think the idea of this app is great. Giving MSM, you know, PrEP information and PEP [post-exposure prophylaxis] information and easier ways to access different locations and information is great, you know, because a lot of us aren't as educated on these topics and aren't—don't have these resources. And using the Internet, it's way harder to access the information just by looking it up on Bing or Google than, you know, the app just gives it to you, you know?...It definitely gives you a lot of information that you*

> *wouldn't have access to otherwise.* [Afro-Latinx, bisexual, age 20 years, the Bronx]

Participants identified a number of features that they believed would be most useful, including the ability to order free HIV or STI test kits, condoms, and lubricant; being able to create personalized HIV testing plans with the location finder; and being able to ask questions to a health professional through the app:

> *I would say that this is the most immediately pertinent thing for me that I've seen in the app so far. I think that it's really important to have a plan when it comes to getting tested and to know, like have a schedule and things like that.* [Specific identity unknown]

> *The map features are working well here...when it's full screen I really like it. I love using maps for everything in my life. And I love seeing like where I am and where I could go.* [White, queer, age 22 years, Boston]

Suggestions for User Interface

Many participants liked the *look and feel* of the app, noting that it was nondescript enough to sufficiently maintain privacy:

> *You don't want it [the app] to be something that someone just scrolling through your homepages will be like, "Oh, that's what that is." [an HIV prevention app] It [the app] doesn't really show that particularly.* [White, gay, age 24 years, Boston]

Participants appreciated the wide range of media types, including colorful icons, GIFs, videos, and text. One participant described a GIF related to accessing PrEP:

> *I feel like they're great animations to, you know—it simplifies what it actually is. It's showing you what you're doing. Like, the individual, it shows you coming from your house, going to the doctor's office and then, you know, talking to the doctor.* [Afro-Latinx/Asian, age 23 years, the Bronx]

However, others expressed that the colors and format made the app appear *basic*:

> *I think it looks sort of basic and plain. Particularly maybe just because of the white background...It makes it look a bit sort of less mature like this and less formal. Which is perhaps not the vibe I'd look for in a health app.* [Asian, gay, age 21 years, Boston]

Some participants suggested that portions of the app felt repetitive and could be better streamlined and organized:

> *I think there are too many icons. Because I tapped the "my activity" button and that led me to ordering condoms. So I think if the choices were more simplified, I think I would understand a little bit more about what I can do with the app.* [White, gay, age 21 years, Boston]

Conversely, others felt that obtaining similar information in a variety of different ways was helpful and a strength of the MyChoices app:

> *I think that it's good to kind of have like a no wrong door approach to it where you can get to it in a variety of ways.* [White, gay, age 20 years, Boston]

Feedback on Language

Some participants felt that the language used was appropriate and made them trust the source:

> *And, yeah, I just think overall like it was actually really medically-driven too which I like that. It kept it very professional.* [Latinx, pansexual, age 18 years, the Bronx]

However, others felt that the language used to present the information was too academic and included too much science jargon:

> *Just, I have a part about comfortability. I think I want to feel comfortable when I'm using it...This is just something friendly, it's supposed to be a guide or resource. It's not supposed to scare me or freak me out.* [Black, queer, age 19 years, the Bronx]

Participants appreciated the places in the app that allowed them to personalize the language so that they could decide what type of language—more or less direct, more or less casual—suited them personally. They also felt that personalization of the language also protects against privacy and confidentiality concerns because you "can put whatever you want":

> *But the on screen notification is cool. I mean, it's gonna be more options. It depends. Everybody's different, so everybody has a different preference to what they, how they like to be notified.* [Afro-Latinx/Asian, age 23 years, the Bronx]

Suggestions for Content and Features

In addition to feedback on the current MyChoices prototype, participants provided suggestions on additional content and functions that might enhance the acceptability and utility of the app. One participant suggested including a section on how to talk to your partner about PrEP:

> *I think that another resource that would be helpful on this page would also be talking to your partners about PrEP...If you're having sex with multiple partners regularly then it might something to be like how to tell your partner that you are on PrEP. Or how to possibly suggest to a partner that going on PrEP might be a good option.* [White, gay, age 24 years, Boston]

Participants also felt that getting tested for HIV, and even talking about HIV, can produce a lot of anxiety and that the app could include more information about what to do if you do get a reactive test to assuage some of those fears:

> *...there should be something in there, like a section that should say, "Oh, if you do have it [HIV]," resources about that, about going to get help, like, there would be something that could help you get that.* [Afro-Latinx, gay, age 20 years, the Bronx]

Similarly, participants described their belief that many young people were still misinformed and lacked adequate knowledge

about HIV. As such, they suggested including more basic information about HIV in the app:

> *...If we want to make this really accessible and make sexual health like an accessible topic, you sort of need to go back to basics, and be like, "Here is what HIV is, here's what happens, here's what it is not, here's how to treat it." And that sort of basic information, I think would really helpful to keep people going back to, "This taught me a lot," you know.* [White, gay, age 19 years, Boston]

In addition to new content, participants included suggestions for new features. For example, we described a potential feature that would use geofencing to notify a user when they are near an HIV testing location and they are *due* for an HIV test according to their created test plan. Participants were excited about this option, saying "that's pretty cool" and "that's a cool feature."

Participants also suggested additional tools for interacting with health professionals, including allowing HIV testing sites to provide results directly through the app:

> *Part of me wished that if – so let's say, you got tested from [Health Center] and there was a way [Health Center] could coordinate so that your results just like pop-up in the app, like you don't have to put it manually and that is like your way of receiving them, too. So it is all in one place and you don't do it manually...And they just like get a push notification like, "Your results are in!" And that's like – you go in and you just like find out that way.* [White, gay, age 21 years, Boston]

Participants also appreciated information about postexposure prophylaxis (PEP) but felt that the need for immediate action warrants easier access to this information:

> *It would've been great if like within this app there is like an, "I've just had unprotected sex. What do I do?" You know, because I think for it to come up only in the part about PEP is like, you know – or for it to be like the check-in and then, oh, in the past 72 hours, what if it has been like more than that 90 hours or something, because I didn't like really, you know, have that time to like be checking a quiz.* [Black, gay, age 20 years, Boston]

Summary of Changes Made to MyChoices After Theater Testing

As noted earlier, the participants saw the value in the content and functionalities of the MyChoices app. They believed that the suggestions that came from brief quizzes were useful, that the testing plan would be helpful to encourage regular testing, that the PrEP information was instructive, and that being able to order HIV or STI test kits and safer sex supplies would be beneficial. However, participants also provided suggestions on how to build on and improve some of these components. As a result of these suggestions, before initiating the technical pilot, we refined some of the language in the app, updated GIFs and icons, and streamlined the flow through the app. Moreover, we expanded app functionalities to include (1) a geolocator function that pings individuals when they are near a testing site and due for HIV testing based on their personalized testing plan, (2) the *Need PEP?!* button that is available at the bottom of every screen on the app to directly connect participants with information about PEP and locations where it is available, (3) additional videos to demonstrate how to use the home testing kits, and (4) emails that are sent to users after downloading the app to introduce them to key features that they may have otherwise missed. Some suggestions made by participants were unable to be implemented into the app, although they were noted as potential ways to enhance future iterations, including receiving test results from clinics through the app, adding a real-time chat feature, syncing reminders with phone calendars, and being able to schedule HIV testing or PrEP appointments through the app.

Findings From the Technical Pilot

The open pilot enrolled 11 participants (Boston, n=6; the Bronx, n=5), and retention at the 2-month follow-up was 100%. Participants' ages ranged from 15 years to 23 years, with a median age of 19 years. YMSM of color made up 91% (10/11) of the sample (Black, non-Hispanic, n=5; Hispanic or Latinx, n=5; Table 2).

Table 2. Baseline demographic characteristics of participants in the technical pilot of a novel HIV prevention app for HIV prevention in young men who have sex with men in in Boston, Massachusetts, and the Bronx, New York City, 2019 (N=11).

Characteristics	Values
Age (years), mean (SD)	18.8 (2.7)
Number of condomless anal sex acts in the past 6 months, mean (SD)	4.0 (6.8)
Study site, n (%)	
Boston, Massachusetts	6 (55)
The Bronx, New York City	5 (45)
Hispanic or Latinx, n (%)	4 (36)
Race, n (%)	
Black	7 (64)
White	3 (27)
Multiracial or other	1 (9)
Sexual orientation, n (%)	
Gay or homosexual	7 (64)
Bisexual	4 (36)
Same gender loving	2 (18)
Queer	1 (9)
Currently in school, n (%)	8 (73)
Highest level of education completed, n (%)	
Less than high school	3 (27)
Some college, or technical or vocational school	6 (55)
4-year college graduate or more	2 (18)
Currently has health insurance, n (%)	11 (100)
Currently has primary care provider, n (%)	9 (82)
HIV test, n (%)	
In the past 3 months	0 (0)
Ever	3 (27)
STI[a] test, n (%)	
In the past 3 months	0 (0)
Ever	6 (55)
Heard of PrEP[b] before the study, n (%)	9 (82)
Discussed PrEP with a health care provider before the study, n (%)	2 (18)
Interested in taking PrEP, n (%)	
Somewhat or very or extremely interested	8 (73)
A little interested	3 (27)

[a]STI: sexually transmitted infection.

[b]PrEP: pre-exposure prophylaxis.

App Feasibility

Participants in the technical pilot accessed the app, on average, 8 times (SD 5.0; range 2-18) over 2 months, with an average duration of 28 minutes per session (SD 38.9). Across all participants, the cumulative time spent in the app ranged from 1.8 minutes to 20.5 hours, with an average of 4 hours and 39 minutes (SD 7 hours). All participants used the test plan feature (11/11, 100%; average number of accesses to this feature 20.2, SD 18.2; range 1-63), and nearly all participants (10/11, 91%) used MyChoices to order HIV or STI self-testing kits and safer sex supplies (average number of accesses 22.6, SD 29.8; range 0-85). Most participants (7/11, 64%) used the app to locate HIV or STI testing centers or PrEP providers (average number of accesses 5.3, SD 5.8; range 0-17), and 4 participants (4/11, 36%) accessed the FAQ feature of the app to access information about

HIV prevention and PrEP (average number of accesses 1.5, SD 3.0; range 0-9).

App Acceptability

At the 2-month follow-up, the mean SUS (0-100) score was 71 (SD 11.8), which is considered above average. Almost all participants (9/11, 82%) agreed that MyChoices was useful, 73% (8/11) were very satisfied with MyChoices, and 91% (10/11) said that they would recommend it to a friend who needed help with getting HIV tests or accessing PrEP. Nearly all participants (9/11, 82%) reported that they would be very likely (2/11, 18%), likely (4/11, 36%), or somewhat likely (3/11, 27%) to use the MyChoices app if it were to become publicly available. The utility of MyChoices for HIV testing and PrEP was also highly rated (Table 3).

Table 3. Utility of the MyChoices app for HIV prevention in young men who have sex with men in Boston, Massachusetts, and the Bronx, New York City, 2019 (N=11).

Dimension	Values[a], n (%)
MyChoices motivated me to get tested for HIV	10 (91)
MyChoices helped me understand whether PrEP[b] would be a good fit for me	9 (82)
MyChoices assisted me in getting tested for HIV	8 (73)
MyChoices helped me understand my risk for getting HIV	7 (64)
MyChoices assisted me in getting started on PrEP	5 (45)
MyChoices motivated me to get on PrEP	5 (45)

[a]Number (percentage) of participants in the technical pilot who indicated that they strongly agreed or agreed with each of the statements.

[b]PrEP: pre-exposure prophylaxis.

The most highly rated features of the MyChoices app in terms of helpfulness were the ability to order self-testing kits and safer sex supplies (10/11, 91%), PrEP information (10/11, 91%), HIV or STI and PrEP locator (9/11, 82%), and personalized testing plans (8/11, 73%). Less helpful features were the check-in quizzes (6/11, 55%) and testing reminders (5/11, 45%). Finally, participants agreed that the MyChoices app had a beneficial impact on their lives in the following ways: getting tested and knowing their HIV status (8/11, 73%) improved their understanding of the risk of HIV (7/11, 64%), feeling good about helping others or community (6/11, 55%), getting access to medical care (6/11, 55%), receiving assistance for getting on PrEP (5/11, 45%), and improving personal relationships (4/11, 36%).

Exit Interviews

All 11 participants completed follow-up exit interviews. Overall, there was a high level of acceptability of the content, interface, and features. Participants commented on the relevant information presented in the app, one participant noting that "[they] learned a lot about PEP and PrEP" and noted that the app facilitated their scheduling a PrEP counseling appointment with their primary care provider. Another participant noted that they "didn't know anything about [PrEP] until [they] used the app." Furthermore, the feature to schedule PrEP appointments "was the best part for [them]" because "it shows you, like, you are and, like, the different centers, testing centers around you, the different communities around you." The participant endorsed how the app helped him feel more connected to the community where they lived. Another participant highlighted how the app provided him with helpful information about health clinics in his area, of which he was unaware despite living in the neighborhood "a good portion of [his] life."

Participants also provided suggestions to improve the app and provided specific feedback on additional resources and features that they found interesting and potentially helpful. One participant suggested including a *myth busting* section in the app to offset false information spread on the web about sexual health. A few participants also noted that some of the information in the app remains repetitive. One participant suggested defining PrEP differently in different areas of the app to reduce repetition and to mitigate a potential lack of understanding of health information.

Discussion

Principal Findings

This study describes the findings from two phases of formative data collection (theater testing of the MyChoices mobile app prototype and the technical pilot of MyChoices fully functional app) to create a mobile app to increase HIV testing and PrEP uptake among young MSM in the United States. HIV prevention apps are proliferating; however, using evidence-based methods for intervention development, including community-centered approaches with iterative feedback from the community, is essential to maximize their impact and reach [35,49], and few HIV prevention apps for youth have been developed and tested using the approaches described in this paper.

In the context of previously reported findings about preferences for mobile HIV prevention apps in young people, our findings support and extend previous reports by providing a sample including a higher proportion of teenage participants, confirming some aspects of previous reports, and documenting some novel findings. Like others [50-52], we found that youth valued having an app that presents credible information relevant to their health and interests in direct and understandable language and the value of having educational information available. Notably,

other qualitative studies of preferences for HIV prevention or HIV management eHealth tools have identified some themes that did not emerge from the youth who participated in our study. For example, other researchers have found youth preferences for features that allow them to interact with other youth [53], and young MSM in other studies reported concerns about privacy and confidentiality of data [54]. Young MSM have also raised questions about what the scope of sexual health apps should be, with some suggesting that related health issues, such as substance use, could be included in a mobile app resource [52]. Our participants also provided suggestions for developing geospatial tools to guide users to prevention services. These issues, which have been rarely reported in other studies, may have emerged from our participants because geospatial tools are becoming more refined and younger people have grown up relying on smartphones as a primary source of navigation. Due to the younger age of our participants compared with many previous studies and because our data were collected more recently, youth expectations around enhanced navigation and geolocation services may be seen as emerging expectations for mobile prevention apps.

Theater testing revealed high levels of interest in the content provided through MyChoices, with youth indicating that the availability of this type of information is both lacking and necessary. Although youth have access to huge amounts of health information through the internet and report using the internet frequently to access this type of information [55,56], they do not always know what to trust [57,58]. In theater testing, youth indicated that one of the major strengths of the app was to have access to a wide range of sexual health information in one place, in multiple formats, that they knew they could trust. In addition, the app's usability score is comparable with the HealthMindr app on which MyChoices was based, although it had a higher proportion of people who would recommend it to friends [33]. This may suggest a successful adaptation for YMSM.

In addition to the content, the features of the MyChoices app were also broadly seen as favorable and useful. In both theater testing and the technical pilot, the most popular features included ordering free HIV and STI self-testing kits, condoms, and lube; ability to search for nearby HIV testing and PrEP care sites; and PrEP information provided in multiple formats (ie, text, videos, GIFs, and infographics). This suggests that YMSM are open to accessing multiple means of HIV prevention support and that providing a large toolbox of HIV prevention options using diverse modalities is essential for reaching this group at the highest risk for HIV acquisition [59].

Although there was consensus on the importance of the content and utility of the features of MyChoices, there was a wide range of views on preferences for the user interface. Some participants appreciated that the language and presentation of information was more *formal*, as it was viewed as more trustworthy and legitimate. However, others felt that the interface was bland and potentially even anxiety producing and that more casual language would render the app more relatable. This dichotomy highlights the ongoing difficulty in developing an app that aims to reach large populations; although YMSM in the United States are a subpopulation of a larger group, they are not a monolith and will have diverse needs and preferences [60-62]. In addition, with technology constantly changing and advancing, mobile apps must be flexible and responsive to these changes in technology and end user preferences [63,64].

Limitations

These results should be interpreted in light of the following limitations. First, the technical pilot was small and used a nonrandomized design, and as a result, it was not powered or designed to evaluate efficacy. Sample sizes for technical pilots are often determined based on practical considerations rather than inferential statistical power calculations [65]. Still, the evaluation of feasibility and acceptability of behavioral interventions in open pilot studies is an important part of the intervention development process and helps inform subsequent randomized controlled trials (RCTs) to evaluate the efficacy of the intervention [66]. Second, social desirability bias may have led participants in the theater testing and technical pilot to speak more positively about their experience of the app during focus groups and exit interviews. To minimize the potential for these biases, participants were continuously reminded that there were no right or wrong answers and that it was important to provide honest responses. Third, for both phases of the study, individuals were enrolled in the Bronx, New York City, and Boston, Massachusetts, only, potentially limiting the generalizability of the findings. Future studies should expand to other regions of the country and outside of large metropolitan areas, particularly in the South, where the HIV epidemic is spreading most rapidly among young Black MSM. Finally, in both phases of the study, participants had to report sexual risk for HIV, no current PrEP use, and no recent HIV testing. This allowed us to ensure that we received input from those at the highest risk and from those who may benefit the most from the app; however, it also limits the generalizability of our findings to less risky populations.

Conclusions

HIV incidence in the United States remains disproportionately high among YMSM, compared with other risk and demographic groups. YMSM are also less likely than their adult peers to know that they are infected with HIV, highlighting an imminent need to increase routine HIV testing and expand access to HIV prevention interventions, including PrEP. Smartphone use is ubiquitous in the United States, and mobile apps offer an opportunity to reach YMSM "where they're at." mHealth apps have proliferated in recent years; however, only a limited number have been theory-driven and developed using evidence-based methods for intervention development. These data from theater testing and a technical pilot show the initial promise, feasibility, and acceptability of the MyChoices app to improve HIV testing and PrEP uptake among YMSM in the United States. The next step involves further pilot testing using an RCT design to determine more accurate effect size estimates; a full-scale RCT efficacy trial; and ultimately, if efficacious, an implementation study to ensure it is disseminated in such a way that maximizes its reach and utility. At each step, iterative evaluation and refinement based on the reflections and experiences of YMSM will be prioritized.

Acknowledgments

This study was supported by the National Institutes of Health Adolescent Trials Network (U01HD068040; principal investigators [PI]: LH and PS), Emory Center for AIDS Research (P30AI050409; PI: C del Rio) for Carekit support, and MAC AIDS Fund. The funding sources played no role in writing or influencing the decision to publish the manuscript.

Conflicts of Interest

None declared.

References

1. HIV Surveillance Report, 2018 (Updated Edition). Centers for Disease Control and Prevention. 2020 May. URL: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2020-06-05] [WebCite Cache ID http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html]
2. Zanoni BC, Mayer KH. The Adolescent and Young Adult HIV Cascade of Care in the United States: Exaggerated Health Disparities. AIDS Patient Care and STDs 2014 Mar;28(3):128-135. [doi: 10.1089/apc.2013.0345]
3. Wejnert C, Hess KL, Rose CE, Balaji A, Smith JC, Paz-Bailey G. Age-Specific Race and Ethnicity Disparities in HIV Infection and Awareness Among Men Who Have Sex With Men—20 US Cities, 2008–2014. J Infect Dis 2015 Oct 20;213(5):776-783. [doi: 10.1093/infdis/jiv500]
4. Kapogiannis B, Koenig L, Xu J, Mayer K, Loeb J, Greenberg L, Adolescent Medicine Trials Network for HIV/AIDS Interventions. The HIV Continuum of Care for Adolescents and Young Adults Attending 13 Urban US HIV Care Centers of the NICHD-ATN-CDC-HRSA SMILE Collaborative. J Acquir Immune Defic Syndr 2020 May 01;84(1):92-100. [doi: 10.1097/QAI.0000000000002308] [Medline: 32267659]
5. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data: United States and 6 dependent areas, 2017. Centers for Disease Control and Prevention. 2019 Jun. URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-24-3.pdf [accessed 2020-06-05]
6. Li Z, Purcell DW, Sansom SL, Hayes D, Hall HI. HIV Transmission Along the Continuum of Care — United States, 2016. MMWR Morb. Mortal. Wkly. Rep 2019 Mar 22;68(11):267-272. [doi: 10.15585/mmwr.mm6811e1]
7. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, iPrEx Study Team. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599 [FREE Full text] [doi: 10.1056/NEJMoa1011205] [Medline: 21091279]
8. Owens DK, Davidson KW, Krist AH, Barry MJ, Cabana M, Caughey AB, et al. Preexposure Prophylaxis for the Prevention of HIV Infection. JAMA 2019 Jun 11;321(22):2203. [doi: 10.1001/jama.2019.6390]
9. US Preventive Services Task Force, Owens DK, Davidson KW, Krist AH, Barry MJ, Cabana M, et al. Screening for HIV Infection: US Preventive Services Task Force Recommendation Statement. JAMA 2019 Jun 18;321(23):2326-2336. [doi: 10.1001/jama.2019.6587] [Medline: 31184701]
10. Phillips A, Cambiano V, Miners A, Lampe F, Rodger A, Nakagawa F, et al. Potential impact on HIV incidence of higher HIV testing rates and earlier antiretroviral therapy initiation in MSM. AIDS 2015 Sep 10;29(14):1855-1862. [doi: 10.1097/QAD.0000000000000767] [Medline: 26372391]
11. MacFadden DR, Tan DH, Mishra S. Optimizing HIV pre-exposure prophylaxis implementation among men who have sex with men in a large urban centre: a dynamic modelling study. J Int AIDS Soc 2016 Sep 23;19(1):20791. [doi: 10.7448/IAS.19.1.20791] [Medline: 27665722]
12. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, HPTN 052 Study Team. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med 2011 Aug 11;365(6):493-505. [doi: 10.1056/NEJMoa1105243] [Medline: 21767103]
13. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. The Lancet 2016 Jan;387(10013):53-60. [doi: 10.1016/S0140-6736(15)00056-2]
14. DiNenno EA, Prejean J, Irwin K, Delaney KP, Bowles K, Martin T, et al. Recommendations for HIV Screening of Gay, Bisexual, and Other Men Who Have Sex with Men — United States, 2017. MMWR Morb. Mortal. Wkly. Rep 2017 Aug 11;66(31):830-832. [doi: 10.15585/mmwr.mm6631a3]
15. Sanchez TH, Sineath RC, Kahle EM, Tregear SJ, Sullivan PS. The Annual American Men's Internet Survey of Behaviors of Men Who Have Sex With Men in the United States: Protocol and Key Indicators Report 2013. JMIR Public Health Surveill 2015 Apr 17;1(1):e3. [doi: 10.2196/publichealth.4314] [Medline: 27227126]
16. Krakower DS, Mimiaga MJ, Rosenberger JG, Novak DS, Mitty JA, White JM, et al. Limited Awareness and Low Immediate Uptake of Pre-Exposure Prophylaxis among Men Who Have Sex with Men Using an Internet Social Networking Site. PLoS One 2012 Mar 28;7(3):e33119 [FREE Full text] [doi: 10.1371/journal.pone.0033119] [Medline: 22470438]

17. Bauermeister J, Meanley S, Pingel E, Soler J, Harper G. PrEP awareness and perceived barriers among single young men who have sex with men. Curr HIV Res 2013 Oct 31;11(7):520-527 [FREE Full text] [doi: 10.2174/1570162x12666140129100411] [Medline: 24476355]
18. Shover CL, Javanbakht M, Shoptaw S, Bolan RK, Lee S, Parsons JT, et al. HIV Preexposure Prophylaxis Initiation at a Large Community Clinic: Differences Between Eligibility, Awareness, and Uptake. Am J Public Health 2018 Oct;108(10):1408-1417. [doi: 10.2105/ajph.2018.304623]
19. Huang YA, Zhu W, Smith DK, Harris N, Hoover KW. HIV Preexposure Prophylaxis, by Race and Ethnicity - United States, 2014-2016. MMWR Morb Mortal Wkly Rep 2018 Oct 19;67(41):1147-1150 [FREE Full text] [doi: 10.15585/mmwr.mm6741a3] [Medline: 30335734]
20. Coy KC, Hazen RJ, Kirkham HS, Delpino A, Siegler AJ. Persistence on HIV preexposure prophylaxis medication over a 2-year period among a national sample of 7148 PrEP users, United States, 2015 to 2017. J Int AIDS Soc 2019 Feb 18;22(2):e25252 [FREE Full text] [doi: 10.1002/jia2.25252] [Medline: 30775846]
21. Doblecki-Lewis S, Liu A, Feaster D, Cohen S, Elion R, Bacon O, et al. Patterns and Correlates of Participant Retention in a Multi-City Pre-Exposure Prophylaxis Demonstration Project. J Acquir Immune Defic Syndr 2018 Sep 01;79(1):62-69 [FREE Full text] [doi: 10.1097/QAI.0000000000001724] [Medline: 29771790]
22. Serota D, Rosenberg E, Sullivan P, Thorne A, Rolle C, Del Rio C, et al. Pre-exposure Prophylaxis Uptake and Discontinuation Among Young Black Men Who Have Sex With Men in Atlanta, Georgia: A Prospective Cohort Study. Clin Infect Dis 2020 Jul 27;71(3):574-582 [FREE Full text] [doi: 10.1093/cid/ciz894] [Medline: 31499518]
23. Hosek S, Rudy B, Landovitz R, Kapogiannis B, Siberry G, Rutledge B, Adolescent Trials Network (ATN) for HIVAIDS Interventions. An HIV Preexposure Prophylaxis Demonstration Project and Safety Study for Young MSM. J Acquir Immune Defic Syndr 2017 Jan 01;74(1):21-29 [FREE Full text] [doi: 10.1097/QAI.0000000000001179] [Medline: 27632233]
24. Arnett J. Emerging adulthood: The winding road from the late teens through the twenties. New York, New York: Oxford University Press; 2014.
25. Gruber H, Voneche J, editors. The Essential Piaget. New York, New York: Basic Books; 1977.
26. Horvath KJ, Rosser BS, Remafedi G. Sexual Risk Taking Among Young Internet-Using Men Who Have Sex With Men. Am J Public Health 2008 Jun;98(6):1059-1067. [doi: 10.2105/ajph.2007.111070]
27. Teva I, Bermúdez MP, Buela-Casal G. Sexual Sensation Seeking, Social Stress, and Coping Styles as Predictors of HIV/STD Risk Behaviors in Adolescents. Youth & Society 2009 Nov 23;42(2):255-277. [doi: 10.1177/0044118x09353572]
28. Smartphone access nearly ubiquitous among teens, while having a home computer varies by income. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/pi_2018-05-31_teenstech_0-04/ [accessed 2020-06-05]
29. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, Internet and Web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015 Mar 28;12(1):173-190 [FREE Full text] [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
30. Catalani C, Philbrick W, Fraser H, Mechael ,, Israelski DM. Open AIDS J 2013 Sep 20;7(1):17-41 [FREE Full text] [doi: 10.2174/1874613620130812003] [Medline: 24133558]
31. Maloney KM, Bratcher A, Wilkerson R, Sullivan PS. Electronic and other new media technology interventions for HIV care and prevention: a systematic review. J Int AIDS Soc 2020 Jan;23(1):e25439 [FREE Full text] [doi: 10.1002/jia2.25439] [Medline: 31909896]
32. Biello KB, Marrow E, Mimiaga MJ, Sullivan P, Hightow-Weidman L, Mayer KH. A Mobile-Based App (MyChoices) to Increase Uptake of HIV Testing and Pre-Exposure Prophylaxis by Young Men Who Have Sex With Men: Protocol for a Pilot Randomized Controlled Trial. JMIR Res Protoc 2019 Jan 07;8(1):e10694 [FREE Full text] [doi: 10.2196/10694] [Medline: 30617042]
33. Sullivan PS, Driggers R, Stekler JD, Siegler A, Goldenberg T, McDougal SJ, et al. Usability and Acceptability of a Mobile Comprehensive HIV Prevention App for Men Who Have Sex With Men: A Pilot Study. JMIR Mhealth Uhealth 2017 Mar 09;5(3):e26 [FREE Full text] [doi: 10.2196/mhealth.7199] [Medline: 28279949]
34. Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R. Preferences for a Mobile HIV Prevention App for Men Who Have Sex With Men. JMIR Mhealth Uhealth 2014 Oct 29;2(4):e47 [FREE Full text] [doi: 10.2196/mhealth.3745] [Medline: 25355249]
35. Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R. Building a Mobile HIV Prevention App for Men Who Have Sex With Men: An Iterative and Community-Driven Process. JMIR Public Health Surveill 2015 Nov 16;1(2):e18 [FREE Full text] [doi: 10.2196/publichealth.4449] [Medline: 27227136]
36. Bandura A. The anatomy of stages of change. Am J Health Promot 1997;12(1):8-10. [doi: 10.4278/0890-1171-12.1.8] [Medline: 10170438]
37. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr 30;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
38. Hightow-Weidman LB, Muessig K, Rosenberg E, Sanchez T, LeGrand S, Gravens L, et al. University of North Carolina/Emory Center for Innovative Technology (iTech) for Addressing the HIV Epidemic Among Adolescents and

Young Adults in the United States: Protocol and Rationale for Center Development. JMIR Res Protoc 2018 Aug 03;7(8):e10365 [FREE Full text] [doi: 10.2196/10365] [Medline: 30076126]
39. Lee S, Kapogiannis BG, Allison S. Improving the Youth HIV Prevention and Care Continuums: The Adolescent Medicine Trials Network for HIV/AIDS Interventions. JMIR Res Protoc 2019 Mar 26;8(3):e12050 [FREE Full text] [doi: 10.2196/12050] [Medline: 30912750]
40. Biello KB, Hosek S, Drucker MT, Belzer M, Mimiaga MJ, Marrow E, et al. Preferences for Injectable PrEP Among Young U.S. Cisgender Men and Transgender Women and Men Who Have Sex with Men. Arch Sex Behav 2018 Oct 19;47(7):2101-2107. [doi: 10.1007/s10508-017-1049-7] [Medline: 28929260]
41. Kazemi DM, Borsari B, Levine MJ, Shehab M, Nelson M, Dooley B, et al. Real-time demonstration of a mHealth app designed to reduce college students hazardous drinking. Psychol Serv 2019 May;16(2):255-259 [FREE Full text] [doi: 10.1037/ser0000310] [Medline: 30407059]
42. Making Health Communication Programs Work : a Planner's Guide.: U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, [Office of Cancer Communications, National Cancer Institute]; 2019. URL: https://www.cancer.gov/publications/health-communication/pink-book.pdf [accessed 2020-06-05]
43. Hightow-Weidman LB, Muessig K, Rosenberg E, Sanchez T, LeGrand S, Gravens L, et al. University of North Carolina/Emory Center for Innovative Technology (iTech) for Addressing the HIV Epidemic Among Adolescents and Young Adults in the United States: Protocol and Rationale for Center Development. JMIR Res Protoc 2018 Aug 03;7(8):e10365 [FREE Full text] [doi: 10.2196/10365] [Medline: 30076126]
44. Ryan GW, Bernard HR. Techniques to Identify Themes. Field Methods 2016 Jul 24;15(1):85-109. [doi: 10.1177/1525822X02239569]
45. Johnson BD, Dunlap E, Benoit E. Organizing "mountains of words" for data analysis, both qualitative and quantitative. Subst Use Misuse 2010 Apr 11;45(5):648-670 [FREE Full text] [doi: 10.3109/10826081003594757] [Medline: 20222777]
46. Brooke J. SUS-A quick and dirty usability scale. In: Jordan P, Thomas B, McClelland I, Weerdmeester B, editors. Usability evaluation in industry. London: CRC Press; 1996:4-7.
47. System Usability Scale (SUS). U.S. General Services Administration. 2020. URL: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html [accessed 2020-06-05]
48. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of usability studies 2009;4(3):114-123. [doi: 10.5555/2835587.2835589]
49. Bailey J, Mann S, Wayal S, Abraham C, Murray E. Digital media interventions for sexual health promotion-opportunities and challenges: a great way to reach people, particularly those at increased risk of sexual ill health. BMJ 2015 Mar 03;350(mar03 8):h1099-h1099. [doi: 10.1136/bmj.h1099] [Medline: 25736806]
50. Fontenot HB, Rosenberger JG, McNair KT, Mayer KH, Zimet G. Perspectives and preferences for a mobile health tool designed to facilitate HPV vaccination among young men who have sex with men. Hum Vaccin Immunother 2019 Jan 30;15(7-8):1815-1823 [FREE Full text] [doi: 10.1080/21645515.2019.1568156] [Medline: 30625049]
51. Saberi P, Siedle-Khan R, Sheon N, Lightfoot M. The Use of Mobile Health Applications Among Youth and Young Adults Living with HIV: Focus Group Findings. AIDS Patient Care STDS 2016 Jun;30(6):254-260. [doi: 10.1089/apc.2016.0044] [Medline: 27214751]
52. Muessig KE, Pike EC, Fowler B, LeGrand S, Parsons JT, Bull SS, et al. Putting prevention in their pockets: developing mobile phone-based HIV interventions for black men who have sex with men. AIDS Patient Care STDS 2013 Apr;27(4):211-222. [doi: 10.1089/apc.2012.0404] [Medline: 23565925]
53. Mulawa MI, LeGrand S, Hightow-Weidman LB. eHealth to Enhance Treatment Adherence Among Youth Living with HIV. Curr HIV/AIDS Rep 2018 Aug 29;15(4):336-349 [FREE Full text] [doi: 10.1007/s11904-018-0407-y] [Medline: 29959649]
54. Holloway IW, Winder TJ, Lea CH, Tan D, Boyd D, Novak D. Technology Use and Preferences for Mobile Phone-Based HIV Prevention and Treatment Among Black Young Men Who Have Sex With Men: Exploratory Research. JMIR Mhealth Uhealth 2017 Apr 13;5(4):e46. [doi: 10.2196/mhealth.6436] [Medline: 28408360]
55. Mitchell KJ, Ybarra ML, Korchmaros JD, Kosciw JG. Accessing sexual health information online: use, motivations and consequences for youth with different sexual orientations. Health Educ Res 2014 Feb 16;29(1):147-157 [FREE Full text] [doi: 10.1093/her/cyt071] [Medline: 23861481]
56. Wartella E, Rideout V, Montague H, Beaudoin-Ryan L, Lauricella A. Teens, Health and Technology: A National Survey. MaC 2016 Jun 16;4(3):13-23. [doi: 10.17645/mac.v4i3.515]
57. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. Soc Sci Med 2005 Apr;60(7):1467-1478. [doi: 10.1016/j.socscimed.2004.08.010] [Medline: 15652680]
58. Sbaffi L, Rowley J. Trust and Credibility in Web-Based Health Information: A Review and Agenda for Future Research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
59. Hosek S, Pettifor A. HIV Prevention Interventions for Adolescents. Curr HIV/AIDS Rep 2019 Feb 1;16(1):120-128 [FREE Full text] [doi: 10.1007/s11904-019-00431-y] [Medline: 30707399]

60. Wilson PA, Meyer IH, Antebi-Gruszka N, Boone MR, Cook SH, Cherenack EM. Profiles of Resilience and Psychosocial Outcomes among Young Black Gay and Bisexual Men. Am J Community Psychol 2016 Mar 28;57(1-2):144-157. [doi: 10.1002/ajcp.12018] [Medline: 27217318]
61. Vincent W, Peterson JL, Storholm ED, Huebner DM, Neilands TB, Calabrese SK, et al. A Person-Centered Approach to HIV-Related Protective and Risk Factors for Young Black Men Who Have Sex with Men: Implications for Pre-exposure Prophylaxis and HIV Treatment as Prevention. AIDS Behav 2019 Oct 12;23(10):2803-2815 [FREE Full text] [doi: 10.1007/s10461-019-02630-5] [Medline: 31407211]
62. Halkitis PN, Figueroa RP. Sociodemographic characteristics explain differences in unprotected sexual behavior among young HIV-negative gay, bisexual, and other YMSM in New York City. AIDS Patient Care STDS 2013 Mar;27(3):181-190 [FREE Full text] [doi: 10.1089/apc.2012.0415] [Medline: 23442029]
63. Allison S, Bauermeister JA, Bull S, Lightfoot M, Mustanski B, Shegog R, et al. The intersection of youth, technology, and new media with sexual health: moving the research agenda forward. J Adolesc Health 2012 Sep;51(3):207-212 [FREE Full text] [doi: 10.1016/j.jadohealth.2012.06.012] [Medline: 22921129]
64. Grov C, Breslow AS, Newcomb ME, Rosenberger JG, Bauermeister JA. Gay and bisexual men's use of the Internet: research from the 1990s through 2013. J Sex Res 2014 Apr 22;51(4):390-409 [FREE Full text] [doi: 10.1080/00224499.2013.871626] [Medline: 24754360]
65. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011 May;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]
66. Onken LS, Carroll KM, Shoham V, Cuthbert BN, Riddle M. Reenvisioning Clinical Science: Unifying the Discipline to Improve the Public Health. Clin Psychol Sci 2014 Jan 01;2(1):22-34 [FREE Full text] [doi: 10.1177/2167702613497932] [Medline: 25821658]

Abbreviations

FAQ: frequently asked question
GIF: Graphics Interchange Format
IRB: Institutional Review Board
iTech: University of North Carolina/Emory Center for Innovative Technology
mHealth: mobile health
MSM: men who have sex with men
PEP: postexposure prophylaxis
PI: principal investigators
PrEP: pre-exposure prophylaxis
RCT: randomized controlled trial
STI: sexually transmitted infection
SUS: System Usability Scale
YMSM: young men who have sex with men

Edited by R Kukafka; submitted 22.10.20; peer-reviewed by K Ingersoll, J Opoku, C Reis; comments to author 08.12.20; revised version received 11.12.20; accepted 08.02.21; published 24.03.21.

Please cite as:
Biello KB, Hill-Rorie J, Valente PK, Futterman D, Sullivan PS, Hightow-Weidman L, Muessig K, Dormitzer J, Mimiaga MJ, Mayer KH
Development and Evaluation of a Mobile App Designed to Increase HIV Testing and Pre-exposure Prophylaxis Use Among Young Men Who Have Sex With Men in the United States: Open Pilot Trial
J Med Internet Res 2021;23(3):e25107
URL: https://www.jmir.org/2021/3/e25107
doi:10.2196/25107
PMID:33759792

Viewpoint

What Every Reader Should Know About Studies Using Electronic Health Record Data but May Be Afraid to Ask

Isaac S Kohane[1], MD, PhD; Bruce J Aronow[2], PhD; Paul Avillach[1], MD, PhD; Brett K Beaulieu-Jones[1], PhD; Riccardo Bellazzi[3,4], PhD; Robert L Bradford[5], BSc; Gabriel A Brat[1], MD, MPH; Mario Cannataro[6,7], MS; James J Cimino[8], MD; Noelia García-Barrio[9], MS; Nils Gehlenborg[1], PhD; Marzyeh Ghassemi[10], PhD; Alba Gutiérrez-Sacristán[1], PhD; David A Hanauer[11], MS, MD; John H Holmes[12], PhD; Chuan Hong[1], PhD; Jeffrey G Klann[13,14], PhD; Ne Hooi Will Loh[15], MBBS, FRCA, FFICM, EDIC; Yuan Luo[16], PhD; Kenneth D Mandl[17], MPH, MD; Mohamad Daniar[18], MSIS; Jason H Moore[19], PhD; Shawn N Murphy[1,20], MD, PhD; Antoine Neuraz[21,22], MD; Kee Yuan Ngiam[15], MBBS, MRCS, MMed; Gilbert S Omenn[23], MD, PhD; Nathan Palmer[1], PhD; Lav P Patel[24], MS; Miguel Pedrera-Jiménez[9], MS; Piotr Sliz[17], PhD; Andrew M South[25], MS, MD; Amelia Li Min Tan[1,26], BSc, PhD; Deanne M Taylor[27,28], PhD; Bradley W Taylor[29], MS; Carlo Torti[7], MD; Andrew K Vallejos[29], MS; Kavishwar B Wagholikar[13,14], MBBS, PhD; The Consortium For Clinical Characterization Of COVID-19 By EHR (4CE)[30]; Griffin M Weber[1*], MD, PhD; Tianxi Cai[1*], SCD

[1]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States

[2]Biomedical Informatics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH, United States

[3]Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy

[4]ICS Maugeri, Pavia, Italy

[5]North Carolina Translational and Clinical Sciences Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[6]Data Analytics Research Center, University Magna Graecia of Catanzaro, Catanzaro, Italy

[7]Department of Medical and Surgical Sciences, University Magna Graecia of Catanzaro, Catanzaro, Italy

[8]Informatics Institute, University of Alabama at Birmingham, Birmingham, AL, United States

[9]Department of Informatics, 12 de Octubre University Hospital, Madrid, Spain

[10]Department of Computer Science and Medicine, University of Toronto, Toronto, ON, Canada

[11]Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI, United States

[12]Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[13]Department of Medicine, Harvard Medical School, Boston, MA, United States

[14]Laboratory of Computer Science, Massachusetts General Hospital, Boston, MA, United States

[15]National University Health Systems, Singapore, Singapore

[16]Department of Preventive Medicine, Northwestern University, Chicago, IL, United States

[17]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, United States

[18]Clinical Research Informatics, Boston Children's Hospital, Boston, MA, United States

[19]Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, United States

[20]Department of Neurology, Massachusetts General Hospital, Boston, MA, United States

[21]Department of Biomedical Informatics, Necker-Enfant Malades Hospital, Assistance Publique - Hôpitaux de Paris, Paris, France

[22]Centre de Recherche des Cordeliers, INSERM UMRS 1138 Team 22, Université de Paris, Paris, France

[23]Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI, United States

[24]Department of Internal Medicine, Division of Medical Informatics, University of Kansas Medical Center, Kansas City, KS, United States

[25]Section of Nephrology, Department of Pediatrics, Brenner Children's Hospital, Wake Forest School of Medicine, Winston Salem, NC, United States

[26]Department of Biomedical Informatics, National University of Singapore, Singapore, Singapore

[27]Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, United States

[28]Department of Pediatrics, Perelman School of Medicine, The University of Pennsylvania, Philadelphia, PA, United States

[29]Clinical and Translational Science Institute, Medical College of Wisconsin, Milwaukee, WI, United States

[30]See Acknowledgments

[*]these authors contributed equally

Corresponding Author:
Isaac S Kohane, MD, PhD

Department of Biomedical Informatics
Harvard Medical School
10 Shattuck Street
Boston, MA, 02115
United States
Phone: 1 617 432 3226
Email: isaac_kohane@harvard.edu

Abstract

Coincident with the tsunami of COVID-19–related publications, there has been a surge of studies using real-world data, including those obtained from the electronic health record (EHR). Unfortunately, several of these high-profile publications were retracted because of concerns regarding the soundness and quality of the studies and the EHR data they purported to analyze. These retractions highlight that although a small community of EHR informatics experts can readily identify strengths and flaws in EHR-derived studies, many medical editorial teams and otherwise sophisticated medical readers lack the framework to fully critically appraise these studies. In addition, conventional statistical analyses cannot overcome the need for an understanding of the opportunities and limitations of EHR-derived studies. We distill here from the broader informatics literature six key considerations that are crucial for appraising studies utilizing EHR data: data completeness, data collection and handling (eg, transformation), data type (ie, codified, textual), robustness of methods against EHR variability (within and across institutions, countries, and time), transparency of data and analytic code, and the multidisciplinary approach. These considerations will inform researchers, clinicians, and other stakeholders as to the recommended best practices in reviewing manuscripts, grants, and other outputs from EHR-data derived studies, and thereby promote and foster rigor, quality, and reliability of this rapidly growing field.

(J Med Internet Res 2021;23(3):e22219) doi:10.2196/22219

KEYWORDS

COVID-19; electronic health records; real-world data; literature; publishing; quality; data quality; reporting standards; reporting checklist; review; statistics

Introduction

What should researchers and clinicians conclude about the recent high-profile retractions of COVID-19 studies based on electronic health record (EHR) data? It is impressive that two publications involving patients with COVID-19, one in The Lancet [1] and the other in the New England Journal of Medicine [2], were determined to be unsound and were retracted in less than 2 months from publication, as these journals' review processes and quality checks are among the most rigorous in the world. Yet, upon closer inspection by those of us familiar with EHR-based research, there were many flaws to these studies involving data quality issues and a lack of transparency that should have been more readily identified during the peer and editorial review process. This is not to say that in-depth statistical analysis might not have eventually uncovered concerns but rather to point out incongruities and anomalies unique to EHR-based studies that should immediately raise concerns to experienced biomedical informaticians, much like an experienced contractor explaining to a homeowner why a competing bid is too good to be true.

In this viewpoint, we present six key questions that are necessary to consider when appraising EHR-based research, especially for research studies investigating the pandemic:

1. How complete are the data?
2. How were the data collected and handled?
3. What were the specific data types?
4. Did the analysis account for EHR variability?
5. Are the data and analytic code transparent?
6. Was the study appropriately multidisciplinary?

In particular, we focus on general aspects of these questions that are crucial to study and data quality and validity of and interpretability of the results and that are broadly applicable to many stakeholders, including researchers and clinicians, in order to optimize the review of submitted manuscripts, published studies, and grant applications containing preliminary data. These desiderata were compiled by the 96 members of the Consortium for Clinical Characterization of COVID-19 by EHR (4CE)—a self-assembled group of collaborating hospitals focused specifically on studying the clinical course of patients with COVID-19 using EHR-based data—most of whom are biomedical informaticians—across 7 countries. 4CE members were invited to contribute their specific key concerns to a shared checklist. This list was then pared down into a less technical list for a more general audience. We excluded those items that are generally considered to be good biostatistical practices (eg, manual review of sample data sets, detecting and understanding outliers [3,4]) to present EHR-specific concerns to a broad biomedical audience. We also excluded recommendations that are contained within the Reporting of Studies Conducted Using Observational Routinely Collected Health Data (RECORD) statement [5,6], which are not specific to EHR-derived data. Finally, we did not focus on the specific limitations of EHR-derived studies, which have been amply documented [7,8], or on the methods to minimize the impact of these limitations,

as this viewpoint is not focused on reviewing specific methodological options for investigators using EHR-derived data, which has been reviewed in detail previously [9-11]. We acknowledge that there are many other criteria that can inform evaluations of EHR-based studies, but we have purposefully limited this discussion to those issues that are most relevant to a general audience, centered on studies investigating the pandemic.

Data Completeness

There are several statistical tests to query data completeness and methods for incorporating missing data [12,13], but here we describe the reasonable expectations for such completeness with knowledge of current, state-of-the-art EHR usage. A publication that is specific about which data were obtained from the EHR (eg, specific laboratory tests or billing codes) is more credible than a study that simply claims it obtained 100% of the EHR data (as did the two recently retracted publications [1,2]). The range of data types from EHRs is extensive and highly varied; each data type requires its own specific quality control and transformations to standard terminologies. For example, laboratory measurements alone can have as many as hundreds of thousands of local codes at a large health care system such as the Veterans Health Administration. In many cases, these data require some level of manual record review to assure data quality and completeness.

Similarly, if a study reports a deidentification procedure, it must describe the details of said procedure. The goals of the deidentification process determine the nature of the deidentification process and the associated regulatory requirements. For example, US hospitals can meet HIPAA (Health Insurance Portability and Accountability Act) standards [14] if they require obfuscation of the counts of patients with rare clinical presentations below a specified prevalence threshold and if they employ date shifting. Knowledge of these methods is essential to analyzing and interpreting the derived data.

Some data types are represented theoretically in the EHR but in practice are only recorded occasionally. For example, standardized codes for smoking history or a family history of specific diseases exist but their underuse is well known. Thus, one cannot assume that the lack of smoking history codes equates to the patient being a nonsmoker. In such scenarios, one must provide an explicit description of the management of missing/null values. Many data elements, such as a complete pulmonary function test, exist in a fragmented form, scattered across different fields in the EHR, and are difficult to extract reliably. In addition, clinical notes allow clinicians greater qualitative expressivity on some of the above values, like smoking history, where they are documented more frequently but not consistently. The quality criteria for reporting narrative content from clinical notes are further addressed below.

Many clinical states are not represented explicitly in the EHR but can be inferred (often referred to as computational phenotypes). When a publication refers to hyperlipidemia, readers should ask themselves whether the hyperlipidemic phenotype is assessed from one or more lipid laboratory tests, billing diagnostic codes, prescription of lipid-lowering medication, or a combination of the above. It is important to document if only structured codes were used or if the phenotype was defined based on information extracted from clinical notes by using natural language processing (NLP) or manual chart review. Either a table describing these phenotypic methods or a reference to a public set of definitions (eg, Phenotype Knowledgebase, PheKB [15]) or a published algorithm with reported accuracy (as seen, for example, in Zhang et al [16] and Ananthakrishnan et al [17]) can provide transparency and precision to these EHR-driven computational phenotypes. The lack of this transparency should be a warning sign. If onset time or temporal trends of clinical events are used as outcomes, it is important to provide sufficient details on how the data were used to derive these outcomes, how granular time was incorporated (eg, by day, 24-hour period, or hour/minute), and to comment on their accuracy, since EHR data are particularly noisy with regards to capturing the timing of events [18,19].

If one uses EHR data to obtain population estimates (eg, prevalence of a complication per 100,000 patients), then additional information should be provided so that readers can determine which subset of patients from that population a given hospital's EHR can capture. For example, if the EHR captures a patient's hospitalization for heart failure, will the EHR also capture the preceding or subsequent outpatient clinic visits related to that hospitalization? With health maintenance organizations, such as Kaiser Permanente, that is much less of a concern, but many hospitals operate in a patchwork system where the patient's data are spread across multiple heterogeneous EHRs that do not necessarily communicate. In our recent COVID-19 study [20], we found many instances in which patients with COVID-19 were transferred from another hospital; unless that other hospital was part of our consortium, it was impossible to have a complete record of their COVID-19 clinical course. It is also important to recognize that a given EHR may not fully capture the clinical course of certain patients, such as those infected with SAR-CoV-2 who have mild symptoms and are discharged home from the emergency room. In these instances, integration of EHR data with data from other sources (eg, primary care providers' offices or nursing homes) may increase the reliability of analysis, although in practice this is rare and such integration methods have to be well documented. EHR systems may also fail to capture acute events that occur outside of the system, especially in the coded data. Leveraging NLP data from the clinical notes can potentially recover partial information if the patient has follow-up visits within that particular system.

Data Collection and Handling

Often the units of measurement and the codes used for data elements like laboratory tests, medications, and diagnoses are not the same across hospitals and may even differ within the same health care system or change over time. Single analytic concepts (eg, the troponin T test) can balloon into dozens of local codes at each hospital, since these tests may be performed at different diagnostic laboratories, each with its own distinct codes or with different technologies over time. Therefore, they have to be "harmonized," or mapped, to agreed-upon standard terminologies and scales [21]. Even when they are the same,

their meaning can differ based on population or practice differences (eg, which sensitive troponin test is used or which reference range defines a test result being normal, or in children rather than in adults, whose normative values often change across the age range) [7]. In both instances, readers should expect that the specific procedures for harmonization or site-specific semantic alignment are described adequately in the Methods section (or via supplementary materials). A summary of this process can become increasingly complex within the usual confines of a Methods section for multisite and international studies where, by necessity, the site-by-site variability is high.

Data Type

There are large methodological divides and divergent ethical challenges between codified data (eg, discrete laboratory values such as serum glucose) and narrative text (eg, discharge summary) from which characterizations are obtained using NLP. While both data types have their own limitations, methods that incorporate both can greatly improve the sensitivity and/or specificity of the clinical characterizations and phenotyping of a group of patients. For example, signs and symptoms are often not codified discreetly or consistently (eg, not entered into the EHR's Problem List) but are written in the clinical notes. Similarly, outpatient medication documentation in clinical notes does not necessarily represent accurately the medications that the patient is actually taking, but prescriptions entered into the EHR may. Combining both codified and NLP data can substantially improve sensitivity and/or specificity and ideally one should always use this complementarity [22-24]. For example, only about 10% of pregnant women with suicide ideation have related codes and vast majority of the cases are only documented in the notes [25]. However, the ability to extract NLP data and the accuracy of those data may be limited by each institution's informatics infrastructure and expertise as well as local institutional review board (IRB) constraints. Furthermore, NLP application to clinical narrative text is relatively new and more prone to large variability in the quality of the obtained characterizations. Particularly in countries with different languages, the NLP techniques and their performance may vary widely. For this reason, readers should expect a reference to the specific NLP methods used and their performance characteristics on data of the sort that the study collected and analyzed. For example, if someone describes the use of an NLP approach on discharge summaries in intensive care units in Italy, but the provided citation was validated only for use in outpatient notes written in English, readers can be legitimately concerned about the accuracy and validity of the patient characterizations in that study. Furthermore, if a study claims very high accuracy, readers should expect a report (or citation of a report) that shows an expert review of the NLP method validated against a representative sample confirming the claimed performance.

Robustness Against EHR Variability

Beyond any variation in human biology across countries and continents, different styles of practice, and how different reimbursement schemes influence styles of practice and use of EHRs, have a very large impact on the nature of EHR data. Therefore, a multinational study should at least acknowledge these differences as a limitation or explicitly attempt to account for them in the analyses. For example, in COVID-19–related research, it has become increasingly apparent that there is an association between patient race/ethnicity and their risk for acquisition of and complications from COVID-19. However, this association is much less detectable in EHR data, as, for example, it is mostly invisible in data from Europe because several countries forbid collecting self-reported race in the EHR. Even in the United States, the coding of different ethnicities or multiracial identification is not standardized. In addition, some countries have far more comprehensive primary care EHR data sharing, whereas others (like the United States) cannot aggregate data systematically and consistently across major health care centers.

Transparency

In order to ensure patients' rights to privacy, patient-level data can rarely be shared outside an institution. In many EHR-driven studies, the code to extract data from a source EHR can be protected by confidentiality agreements with the EHR vendor and is thus difficult to share. Nonetheless, the code or algorithm for creating the variables used for analyses should be provided even if the detailed data extraction procedures are not shared because of commercial restrictions. Running the code on synthetic data sets that follow a standard data model can demonstrate code functionality and facilitate code reuse [26]. The code used to conduct statistical analyses and create visualizations—after data extraction—should also be shared in public repositories to enable other researchers to follow each step of the analysis and provide further transparency. While there are significant challenges to sharing patient-level data, one can share intermediate results and aggregate distributions to increase transparency and understand between-institution differences [27]. One should archive the data used for analyses, along with the associated data extraction codes, at the local institution to ensure reproducibility. Authors should also make the deidentified data available—either publicly in a repository or by request. While only a small fraction of readers typically look at the code, whether referenced on a file server or shared as supplementary methods, the availability of the code provides reassurance and validation that the study utilized proper methodologies.

Multidisciplinary Approach

There may come a time when data can be aggregated automatically from multiple EHR environments to answer a particular question without relying on a human to understand the particular idiosyncrasies of each institution's data and EHR system. Until that day, effective EHR data set analysis requires collaboration with clinicians and scientists who have knowledge of the diseases being studied and the practices of their particular health care systems; informaticians with experience in the underlying structures of biomedical record repositories at their own institutions and the characteristics of their data; data

harmonization experts to help with data transformation, standardization, integration, and computability; statisticians and epidemiologists well versed in the limitations and opportunities of EHR data sets and related sources of potential bias; machine learning experts; and at least one expert in regulatory and ethical standards. Data provenance records should already exist to ensure compliance with privacy standards, so that authors can readily point to these processes and reference institutional officials who grant data access similarly to IRBs. In our experience, we often have an interdisciplinary team participate in the process of establishing the research question and study design, defining the data elements, and determining what analyses can be performed given the available data. It is also important that people with complementary skills work together to review and interpret the results [28]. Each of these steps is a major contribution deserving of authorship. Just as a population genetics study reporting across countries often has dozens of authors, so do we expect multihospital EHR-driven studies to acknowledge and name the individuals as authors and in doing so provide accountability for the dozens of procedures, checks, and balances necessary for the reliable extraction of EHR patient data. Consequently, contribution statements should list explicitly the responsibilities of each author with regard to study conceptualization and design, data extraction, data harmonization, data integration, data analysis, results interpretation, and regulatory and ethical oversight. Additionally, although reputation is sometimes overvalued, having *no* reputation or at least a track record of appropriate success should trigger greater attention to documenting the process to reach the same level of trust. Unlike a mathematical proof, simple inspection of the data may be insufficient and will become increasingly so in the era of data generated by machine learning algorithms purposefully built for the task of conditioning data to appear real. Trust and accountability become essential companions to transparency and clarity during the EHR analytic process.

Conclusion

Similar to publications from the early days of the genomic revolution, which initially included extensive sections on DNA sequencing validation, methods, reagents, and conditions that became progressively briefer as trust was built and the methods commoditized, comprehensively and transparently reported methods of EHR data extraction and transformation are at least as important as subsequent statistical analysis and interpretation. We need to be open and transparent about the inherent limitations of the data and the analyses. We should also acknowledge alternative interpretations of the results (eg, outlier prescribing practices in one country that confound the apparent effects of that drug in that country). Extra caution is also needed in how we draw causal inferences from EHR data, especially given the noisiness and incompleteness of the data in addition to several sources of bias, though application of a causal model framework and specific causal inference methods may help mitigate some of these concerns. The recommendations we have outlined here (see Table 1 for our 12-item checklist) do not substitute for a durable research infrastructure that would enable tracking EHR data provenance along explicit source, ownership, and data protocols, which would allow for rigorous and routine quality assurance in the use of EHR data [29].

Table 1. 12-item checklist to assess electronic health record (EHR) data–driven studies.

Item	Reassuring	Concerning
Defining study cohort/data extraction	Reporting the precise definition of the domains and/or subsets of EHR data extracted for the study cohort and the information system sources	100% of the EHR said to be extracted or no specification of which subsets of the EHR data were obtained
Deidentification	Specific deidentification algorithm documented with acknowledgment of analytic consequences/limitations	Only a statement that deidentification was performed
Defining clinical variables/data type–specific omissions/limitations	For data types represented poorly in EHR codified data, either NLP[a] is deployed on the EHR clinical notes or additional data sources (eg, self-reported questionnaires) are used. Procedures to deal with missing values should also be made explicit	Referencing data types like family/social history without explaining how they are obtained through NLP or exceptional codified data practice
Phenotypic transparency	Computational phenotypes that are more than just a specific native EHR variable (eg, hyperlipidemia vs a specific LDL[b] measurement) are either defined in the study or a citation is given to algorithmic phenotype definitions	Clinical phenotypes are used in the study without specifying how they were derived from the EHR data
Generalizing EHR findings to the population/population denominator	Study heavily cautions on using prevalence/incidence estimates from the EHR data or refers to empirical estimates on how much of a patient's entire health care is captured in that particular EHR	Direct estimates of prevalence or incidence from EHR frequencies without justifying that generalization
Data collection	Clinical forms or data models implemented in health care information systems are shared or clearly described. This includes the coding systems used	Mention structured data without specifying the clinical forms or data models. Mention coded data without mentioning coding systems
Data transformation/harmonization	Data transformation process shared or clear description of which methods were used to harmonize data to a standardized terminology, scale units, and account for different local usage	Mention of harmonization methods without specifying which ones and what problems were identified and addressed/overcome
Textual vs codified data	If textual data are used in the study, then specification of which clinical notes, in what language, with which NLP algorithm with either an explanation of or a citation to that algorithm's validation, sensitivity, and specificity for comparable data	Harmonization efforts for codified and textual data treated as if they are the same process. Lack of specificity in describing the NLP algorithm and performance
Manual coding of data	Qualifications of coders described, formal coding criteria described or at least mentioned, intercoder reliability measured and reported	No description of process for turning text or nonstandard coded data into standard coded data; use of crowd-sourced coders (eg, graduate students or Mechanical Turk) without mention of quality assurance processes
Regional and global variation	A study describes how they adjust for (or exclude) differences that are due to variation in practice, regulation, and clinical documentation through the EHR from site to site	A study says they adjusted for regional or country differences in practice or EHR documentation but do not describe how they do it
Sharing analytic code	Analytic code is deposited in a public repository or study-specific public website	Code is not shared or only "shared on demand"
Acknowledge a multidisciplinary team	Authorships for all parts of the extraction-through-analysis pipeline with precision as to each contribution	Health care system sources not named or local health care system site collaborators not named

[a]NLP: natural language processing.

[b]LDL: low-density lipoprotein.

Finally, in crises such as the COVID-19 pandemic, we need to recognize that many studies can contribute to our understanding of what is happening to our patients and how our practices might affect patient outcomes. Overly generalized conclusions will likely strain the boundaries of what can be reasonably inferred from the kinds of data currently obtained through EHRs. Recommendations that flow from overly broad claims may irreversibly harm stakeholders, including patients and clinicians. Increased reader awareness of EHR-derived data quality indicators is crucial in critically appraising EHR-driven studies and to prevent harm from misleading studies, which will ensure sustainable quality in this rapidly growing field.

Acknowledgments

The members of the Consortium for Clinical Characterization of COVID-19 By EHR (4CE) are as follows: Adem Albayrak, Danilo F Amendola, Li LLJ Anthony, Bruce J Aronow, Andrew Atz, Paul Avillach, Brett K Beaulieu-Jones, Douglas S Bell,

Antonio Bellasi, Riccardo Bellazzi, Vincent Benoit, Michele Beraghi, José Luis Bernal Sobrino, Mélodie Bernaux, Romain Bey, Alvar Blanco Martínez, Martin Boeker, Clara-Lea Bonzel, John Booth, Silvano Bosari, Florence T Bourgeois, Robert L Bradford, Gabriel A Brat, Stéphane Bréant, Mauro Bucalo, Anita Burgun, Tianxi Cai, Mario Cannataro, Aize Cao, Charlotte Caucheteux, Julien Champ, Luca Chiovato, James J Cimino, Tiago K Colicchio, Sylvie Cormont, Sébastien Cossin, Jean Craig, Juan Luis Cruz Bermúdez, Arianna Dagliati, Mohamad Daniar, Christel Daniel, Anahita Davoudi, Batsal Devkota, Julien Dubiel, Scott L DuVall, Loic Esteve, Shirley Fan, Robert W Follett, Paula SA Gaiolla, Thomas Ganslandt, Noelia García Barrio, Nils Gehlenborg, Alon Geva, Tobias Gradinger, Alexandre Gramfort, Romain Griffier, Nicolas Griffon, Olivier Grisel, Alba Gutiérrez-Sacristán, David A Hanauer, Christian Haverkamp, Martin Hilka, John H Holmes, Chuan Hong, Petar Horki, Meghan R Hutch, Richard Issitt, Anne Sophie Jannot, Vianney Jouhet, Mark S Keller, Katie Kirchoff, Jeffrey G Klann, Isaac S Kohane, Ian D Krantz, Detlef Kraska, Ashok K Krishnamurthy, Sehi L'Yi, Trang T Le, Judith Leblanc, Guillaume Lemaitre, Leslie Lenert, Damien Leprovost, Molei Liu, Ne Hooi Will Loh, Yuan Luo, Kristine E Lynch, Sadiqa Mahmood, Sarah Maidlow, Alberto Malovini, Kenneth D Mandl, Chengsheng Mao, Patricia Martel, Aaron J Masino, Michael E Matheny, Thomas Maulhardt, Michael T McDuffie, Arthur Mensch, Marcos F Minicucci, Bertrand Moal, Jason H Moore, Jeffrey S Morris, Michele Morris, Karyn L Moshal, Sajad Mousavi, Danielle L Mowery, Douglas A Murad, Shawn N Murphy, Kee Yuan Ngiam, Jihad Obeid, Marina P Okoshi, Karen L Olson, Gilbert S Omenn, Nina Orlova, Brian D Ostasiewski, Nathan P Palmer, Nicolas Paris, Lav P Patel, Miguel Pedrera Jimenez, Hans U Prokosch, Robson A Prudente, Rachel B Ramoni, Maryna Raskin, Siegbert Rieg, Gustavo Roig Domínguez, Elisa Salamanca, Malarkodi J Samayamuthu, Arnaud Sandrin, Emily Schiver, Juergen Schuettler, Luigia Scudeller, Neil Sebire, Pablo Serrano Balazote, Patricia Serre, Arnaud Serret-Larmande, Domenick Silvio, Piotr Sliz, Jiyeon Son, Andrew M South, Anastasia Spiridou, Amelia LM Tan, Bryce WQ Tan, Byorn WL Tan, Suzana E Tanni, Deanne M Taylor, Valentina Tibollo, Patric Tippmann, Andrew K Vallejos, Gael Varoquaux, Jill-Jênn Vie, Shyam Visweswaran, Kavishwar B Wagholikar, Lemuel R Waitman, Demian Wassermann, Griffin M Weber, Yuan William, Zongqi Xia, Alberto Zambelli, Aldo Carmona, Charles Sonday, and James Balshi.

Authors' Contributions

ISK led the 4CE international consortium, conceived and designed the study, and drafted the manuscript. TC led 4CE analytics strategies and made contributions to the study design and drafting of the manuscript. JJC contributed a validation strategy and made edits to the manuscript. NG-B was responsible for data extraction and transformation to 4CE format and quality control of the results and made internal contributions. NG led 4CE visualization strategies and made contributions/edits to the manuscript. JGK contributed to the 4CE validation strategy and data submission strategies and made edits to the manuscript. KDM made contributions to the text and framework and made edits to the manuscript. DM was involved in data extraction and transformation to 4CE format. SNM led 4CE data validation strategies and made contributions/edits to the manuscript. GSO made contributions to strategy and edits to the manuscript. NP contributed to 4CE data analysis, aggregation, and quality control. KBW contributed to validation strategies and made edits to the manuscript. BJA, PA, BKB-J, RB, RLB, GAB, MC, MG, AG-S, DAH, JHH, CH, NHW, YL, JHM, AN, KYN, LPP, MP-J, PS, AMS, ALMT, DMT, BMT, CT, AKV, and GMW made contributions/edits to the manuscript.

Conflicts of Interest

RB and AM are shareholders of Biomeris srl. GSO is affiliated with BoD, Galectin Therapeutics, Angion Biomedica, and Amesite, Inc. DMT consulted on a legal matter for AstraZeneca last year.

References

1. Mehra MR, Desai SS, Ruschitzka F, Patel AN. RETRACTED: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. The Lancet 2020 May [FREE Full text] [doi: 10.1016/S0140-6736(20)31180-6]
2. Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN. Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. N Engl J Med 2020 Jun 18;382(25):e102. [doi: 10.1056/nejmoa2007621]
3. Cox D, Donnelly C. Principles of Applied Statistics. Cambridge, UK: Cambridge University Press; 2011.
4. Eriksson L, Byrne T, Johansson E, Trygg J, Vikström C. Multi- and Megavariate Data Analysis Basic Principles and Applications. Malmo, Sweden: Umetrics Academy; 2013.
5. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med 2015 Oct 6;12(10):e1001885 [FREE Full text] [doi: 10.1371/journal.pmed.1001885] [Medline: 26440803]
6. Langan SM, Schmidt SA, Wing K, Ehrenstein V, Nicholls SG, Filion KB, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). BMJ 2018 Nov 14;363:k3532 [FREE Full text] [doi: 10.1136/bmj.k3532] [Medline: 30429167]
7. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, et al. Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research. Medical Care 2013;51:S30-S37. [doi: 10.1097/mlr.0b013e31829b1dbd]

8. Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. J Med Internet Res 2018 May 29;20(5):e185 [FREE Full text] [doi: 10.2196/jmir.9134] [Medline: 29844010]
9. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC) 2016 Sep 11;4(1):1244 [FREE Full text] [doi: 10.13063/2327-9214.1244] [Medline: 27713905]
10. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc 2013 Jan 01;20(1):144-151 [FREE Full text] [doi: 10.1136/amiajnl-2011-000681] [Medline: 22733976]
11. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using Electronic Health Records for Population Health Research: A Review of Methods and Applications. Annu Rev Public Health 2016 Mar 18;37(1):61-81 [FREE Full text] [doi: 10.1146/annurev-publhealth-032315-021353] [Medline: 26667605]
12. Capocaccia R, De Angelis R. Estimating the completeness of prevalence based on cancer registry data. Statist Med 1997 Feb 28;16(4):425-440. [doi: 10.1002/(sici)1097-0258(19970228)16:4<425::aid-sim414>3.0.co;2-z]
13. Smirnov VB. Earthquake catalogs: Evaluation of data completeness. Volc Seis 1998;19:497-510 [FREE Full text]
14. Methods for De-identification of PHI. Office for Civil Rights. 2015 Nov 6. URL: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html [accessed 2020-06-16]
15. Kirby J, Speltz P, Rasmussen L, Basford M, Gottesman O, Peissig P, et al. PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. J Am Med Inform Assoc 2016 Nov;23(6):1046-1052 [FREE Full text] [doi: 10.1093/jamia/ocv202] [Medline: 27026615]
16. Zhang J, Can A, Lai PMR, Mukundan S, Castro VM, Dligach D, et al. Age and morphology of posterior communicating artery aneurysms. Sci Rep 2020 Jul 14;10(1):11545 [FREE Full text] [doi: 10.1038/s41598-020-68276-9] [Medline: 32665589]
17. Ananthakrishnan AN, Cagan A, Cai T, Gainer VS, Shaw SY, Churchill S, et al. Statin Use Is Associated With Reduced Risk of Colorectal Cancer in Patients With Inflammatory Bowel Diseases. Clin Gastroenterol Hepatol 2016 Jul;14(7):973-979 [FREE Full text] [doi: 10.1016/j.cgh.2016.02.017] [Medline: 26905907]
18. Uno H, Ritzwoller DP, Cronin AM, Carroll NM, Hornbrook MC, Hassett MJ. Determining the Time of Cancer Recurrence Using Claims or Electronic Medical Record Data. JCO Clinical Cancer Informatics 2018 Dec(2):1-10. [doi: 10.1200/cci.17.00163]
19. Liu C, Wang F, Hu J, Xiong H. Temporal Phenotyping from Longitudinal Electronic Health Records: A Graph Based Framework. New York, NY: Association for Computing Machinery; 2015 Presented at: KDD '15: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data; August 2015; Sydney, NSW, Australia p. 705-714. [doi: 10.1145/2783258.2783352]
20. Brat G, Weber G, Gehlenborg N, Avillach P, Palmer N, Chiovato L, et al. International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. NPJ Digit Med 2020;3:109 [FREE Full text] [doi: 10.1038/s41746-020-00308-0] [Medline: 32864472]
21. Klann J, Abend A, Raghavan V, Mandl K, Murphy S. Data interchange using i2b2. J Am Med Inform Assoc 2016 Sep;23(5):909-915 [FREE Full text] [doi: 10.1093/jamia/ocv188] [Medline: 26911824]
22. Ananthakrishnan AN, Cai T, Savova G, Cheng S, Chen P, Perez RG, et al. Improving Case Definition of Crohn s Disease and Ulcerative Colitis in Electronic Medical Records Using Natural Language Processing. Inflammatory Bowel Diseases 2013;19(7):1411-1420. [doi: 10.1097/mib.0b013e31828133fd]
23. Ning W, Chan S, Beam A, Yu M, Geva A, Liao K, et al. Feature extraction for phenotyping from semantic and knowledge resources. J Biomed Inform 2019 Mar;91:103122 [FREE Full text] [doi: 10.1016/j.jbi.2019.103122] [Medline: 30738949]
24. Zhang Y, Cai T, Yu S, Cho K, Hong C, Sun J, et al. High-throughput phenotyping with electronic medical record data using a common semi-supervised approach (PheCAP). Nat Protoc 2019 Dec 20;14(12):3426-3444 [FREE Full text] [doi: 10.1038/s41596-019-0227-6] [Medline: 31748751]
25. Zhong Q, Karlson EW, Gelaye B, Finan S, Avillach P, Smoller JW, et al. Screening pregnant women for suicidal behavior in electronic medical records: diagnostic codes vs. clinical notes processed by natural language processing. BMC Med Inform Decis Mak 2018 May 29;18(1):30 [FREE Full text] [doi: 10.1186/s12911-018-0617-7] [Medline: 29843698]
26. Morin A, Urban J, Adams PD, Foster I, Sali A, Baker D, et al. Research priorities. Shining light into black boxes. Science 2012 Apr 13;336(6078):159-160 [FREE Full text] [doi: 10.1126/science.1218263] [Medline: 22499926]
27. Beaulieu-Jones BK, Greene CS. Reproducibility of computational workflows is automated using continuous analysis. Nat Biotechnol 2017 Apr 13;35(4):342-346 [FREE Full text] [doi: 10.1038/nbt.3780] [Medline: 28288103]
28. Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, et al. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. BMJ 2015 Apr 24;350(apr24 11):h1885-h1885 [FREE Full text] [doi: 10.1136/bmj.h1885] [Medline: 25911572]
29. Geissbuhler A, Safran C, Buchan I, Bellazzi R, Labkoff S, Eilenberg K, et al. Trustworthy reuse of health data: a transnational perspective. Int J Med Inform 2013 Jan;82(1):1-9. [doi: 10.1016/j.ijmedinf.2012.11.003] [Medline: 23182430]

Abbreviations

4CE: Consortium for Clinical Characterization of COVID-19 by EHR
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
RECORD: Reporting of Studies Conducted Using Observational Routinely Collected Health Data
NLP: natural language processing
IRB: institutional review board
PheKB: Phenotype Knowledgebase

Edited by R Kukafka; submitted 13.07.20; peer-reviewed by N Delvaux, M Adly, P Harris, A Adly, A Adly, J Li, L Genaro; comments to author 04.08.20; revised version received 14.09.20; accepted 10.01.21; published 02.03.21.

Please cite as:
Kohane IS, Aronow BJ, Avillach P, Beaulieu-Jones BK, Bellazzi R, Bradford RL, Brat GA, Cannataro M, Cimino JJ, García-Barrio N, Gehlenborg N, Ghassemi M, Gutiérrez-Sacristán A, Hanauer DA, Holmes JH, Hong C, Klann JG, Loh NHW, Luo Y, Mandl KD, Daniar M, Moore JH, Murphy SN, Neuraz A, Ngiam KY, Omenn GS, Palmer N, Patel LP, Pedrera-Jiménez M, Sliz P, South AM, Tan ALM, Taylor DM, Taylor BW, Torti C, Vallejos AK, Wagholikar KB, The Consortium For Clinical Characterization Of COVID-19 By EHR (4CE), Weber GM, Cai T
What Every Reader Should Know About Studies Using Electronic Health Record Data but May Be Afraid to Ask
J Med Internet Res 2021;23(3):e22219
URL: https://www.jmir.org/2021/3/e22219
doi:10.2196/22219
PMID:33600347

Original Paper

Underrepresentation of Phenotypic Variability of 16p13.11 Microduplication Syndrome Assessed With an Online Self-Phenotyping Tool (Phenotypr): Cohort Study

Jianqiao Li[1,2*], MSc; Margaret A Hojlo[3,4*], BA; Sampath Chennuri[5], MSc; Nitin Gujral[5], MBA, MSc; Heather L Paterson[1,2], BSc; Kent A Shefchek[6], MSc; Casie A Genetti[1,2], MSc; Emily L Cohn[5], MPH; Kara C Sewalk[7], MPH; Emily A Garvey[3,4], BA; Elizabeth D Buttermore[8], PhD; Nickesha C Anderson[9], PhD; Alan H Beggs[1,2,10], PhD; Pankaj B Agrawal[1,2,10,11], MD, MMSc; John S Brownstein[5,10], PhD; Melissa A Haendel[12], PhD; Ingrid A Holm[1,2,10], MD, MPH; Joseph Gonzalez-Heydrich[2,3,4,13*], MD; Catherine A Brownstein[1,2,4,10*], MPH, PhD

[1]Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, United States
[2]The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA, United States
[3]Department of Psychiatry and Behavioral Sciences, Boston Children's Hospital, Boston, MA, United States
[4]Tommy Fuss Center for Neuropsychiatric Disease Research, Boston Children's Hospital, Boston, MA, United States
[5]Innovation and Digital Health Accelerator, Boston Children’s Hospital, Boston, MA, United States
[6]Department of Environmental and Molecular Toxicology, Oregon State University, Corvallis, OR, United States
[7]Computational Epidemiology Group, Boston Children’s Hospital, Boston, MA, United States
[8]Human Neuron Core, Translational Neuroscience Center, Boston Children’s Hospital, Boston, MA, United States
[9]Department of Neurology, Boston Children’s Hospital, Boston, MA, United States
[10]Department of Pediatrics, Harvard Medical School, Boston, MA, United States
[11]Division of Newborn Medicine, Boston Children’s Hospital and Harvard Medical School, Boston, MA, United States
[12]Center for Health Artificial Intelligence, University of Colorado Anschutz, Aurora, CO, United States
[13]Department of Psychiatry, Harvard Medical School, Boston, MA, United States
[*]these authors contributed equally

Corresponding Author:
Catherine A Brownstein, MPH, PhD
Division of Genetics and Genomics
Boston Children's Hospital
300 Longwood Avenue
Boston, MA, 02115
United States
Phone: 1 6173554764
Email: catherine.brownstein@childrens.harvard.edu

Abstract

Background: 16p13.11 microduplication syndrome has a variable presentation and is characterized primarily by neurodevelopmental and physical phenotypes resulting from copy number variation at chromosome 16p13.11. Given its variability, there may be features that have not yet been reported. The goal of this study was to use a patient “self-phenotyping” survey to collect data directly from patients to further characterize the phenotypes of 16p13.11 microduplication syndrome.

Objective: This study aimed to (1) discover self-identified phenotypes in 16p13.11 microduplication syndrome that have been underrepresented in the scientific literature and (2) demonstrate that self-phenotyping tools are valuable sources of data for the medical and scientific communities.

Methods: As part of a large study to compare and evaluate patient self-phenotyping surveys, an online survey tool, Phenotypr, was developed for patients with rare disorders to self-report phenotypes. Participants with 16p13.11 microduplication syndrome were recruited through the Boston Children's Hospital 16p13.11 Registry. Either the caregiver, parent, or legal guardian of an affected child or the affected person (if aged 18 years or above) completed the survey. Results were securely transferred to a Research Electronic Data Capture database and aggregated for analysis.

Results: A total of 19 participants enrolled in the study. Notably, among the 19 participants, aggression and anxiety were mentioned by 3 (16%) and 4 (21%) participants, respectively, which is an increase over the numbers in previously published literature. Additionally, among the 19 participants, 3 (16%) had asthma and 2 (11%) had other immunological disorders, both of which have not been previously described in the syndrome.

Conclusions: Several phenotypes might be underrepresented in the previous 16p13.11 microduplication literature, and new possible phenotypes have been identified. Whenever possible, patients should continue to be referenced as a source of complete phenotyping data on their condition. Self-phenotyping may lead to a better understanding of the prevalence of phenotypes in genetic disorders and may identify previously unreported phenotypes.

(J Med Internet Res 2021;23(3):e21023) doi:10.2196/21023

KEYWORDS

self-phenotyping; 16p13.11 microduplication syndrome; copy number variation; genetics; incomplete penetrance; phenotype; variable presentation; human phenotype ontology; online survey; digital health

Introduction

16p13.11 microduplication syndrome is a rare chromosome duplication syndrome associated with copy number variation (CNV) at the chromosome 16p13.11 locus. The syndrome has high variability in phenotype. Clinical case reports have shown that patients with 16p13.11 microduplication (dup16p13.11) may experience intellectual disability, speech delay, and emotional and behavioral disorders like attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) [1-4]. Other neurodevelopmental phenotypes related to this CNV include perinatal hypotonia and feeding difficulties, gross motor delay, epilepsy, and schizophrenia [1,3,5]. Abnormal brain magnetic resonance imaging findings have also been described [1-4]. Additional medical conditions related to 16p13.11 microduplication include cardiovascular disease, as well as a range of congenital abnormalities of varying severity including pulmonary stenosis, coarctation of the aorta, thoracic aortic aneurysm dissection, hypermobile joints, hand/foot deformities, microcephaly and macrocephaly, umbilical hernia, and vision problems such as strabismus, myopia, and amblyopia [4,5].

There are considerable challenges in predicting the clinical outcomes of those with 16p13.11 microduplication. For one, there is incomplete penetrance in which both affected and unaffected members from the same family have been found to carry the same CNV, while *de novo* cases also occasionally occur [2-5]. In addition, variable expressivity of the dup16p13.11 phenotypes may occur as a result of the size of the duplication, reportedly ranging from several kilobases to a few megabases. A majority of the known 16p13.11 microduplications include duplication of the gene *NDE1*, which has long been suggested as the primary candidate gene for the neurological and behavioral phenotypes in affected patients [1,4,6-8].

A potentially powerful approach to understanding the complex phenotypic spectrum of 16p13.11 microduplication syndrome is to collect phenotypic data from patients themselves (or their caregivers), as they experience the symptoms and effects of their condition. GenomeConnect, the National Institutes of Health–funded Clinical Genome Resource (ClinGen) patient registry, developed a patient self-phenotyping survey, which asks patient-friendly questions that have been mapped to a set of high-level human phenotype ontology (HPO) terms [9-11]. HPO is a standardized vocabulary of phenotypic abnormalities encountered in human disease, whereby symptoms and characteristic phenotypic findings (a phenotypic profile) are captured using a logically constructed hierarchy of phenotypic terms [12,13]. An alternative method for self-phenotyping is for patients to generate HPO terms for their condition directly. Our group developed a “layperson” HPO survey called “Phenotypr“ to capture patient phenotypes by translating most standard HPO terms into layperson language that would be easy for patients to comprehend and use (eg, a layperson term for “hypotonia” would be “muscle weakness”) [14,15]. We tested the GenomeConnect and Phenotypr surveys computationally and in patients with known rare diseases.

Here, we describe the results of a subset of participants in the larger study who had 16p13.11 microduplication syndrome and completed the Phenotypr survey. The primary aim was to determine if there were self-identified phenotypes that were underrepresented in previous reports of patients with 16p13.11 microduplication syndrome.

Methods

Recruitment

Participants were the caregivers, parents, or legal guardians of individuals with 16p13.11 microduplication or the affected individuals themselves, and were aged 18 years or older. Individuals with 16p13.11 microduplication were recruited through the Boston Children’s Hospital online 16p13.11 Participant Registry. Individuals who had previously joined the Registry and self-reported a 16p13.11 microduplication diagnosis were sent an informational email about the Phenotypr study. In addition, a blurb about the study that included research team contact information was posted on 16p13.11 microduplication Facebook groups, and potential participants contacted us directly. Participants received a US $15 Amazon gift card for completing the Phenotypr survey. Participants provided chromosomal microarray reports to confirm diagnoses. The study was approved by the Boston Children’s Hospital Institutional Review Board.

Phenotypr Survey Development

We developed Phenotypr as a freely available tool that allows patients to record their conditions in layperson and medical HPO terms (see examples of the completion process in Figures 1-3). In Phenotypr, patients first selected the body systems affected by their condition. Seventeen body system options were provided, such as "Growth," "Ears/Hearing," and "Brain/Nervous System." Participants then typed out their symptoms into the symptom search tool, with search result filters applied based on the body systems that they had selected. Tips for entering symptoms were provided, such as reminders to be as specific as possible and include conditions not local to a certain body part (eg, sensitivity to pain). Phenotypr autocompleted each entry with the layperson HPO term, as well as the standard HPO term, and partitioned the terms into anatomically specific sections. The survey ended with a brief demographics form and open-text feedback boxes. Once the survey was complete, the list of the standard HPO terms that corresponded to the layperson HPO terms was downloadable in PDF format.

Phenotypr consisted of a back-end administrative tool for updating ontology versions, user and administrative permissions, and support for alternative implementations; a front-facing public site; and a back-end relational Research Electronic Data Capture (REDCap) database for securely housing the data [16]. Ontology autocomplete features were implemented by processing the HPO and loading structured data into an Apache Solr search engine [17]. The user interface was implemented as a single page application with Vue.js [18].

Figure 1. Data entry for Phenotypr: relevant body system selection.

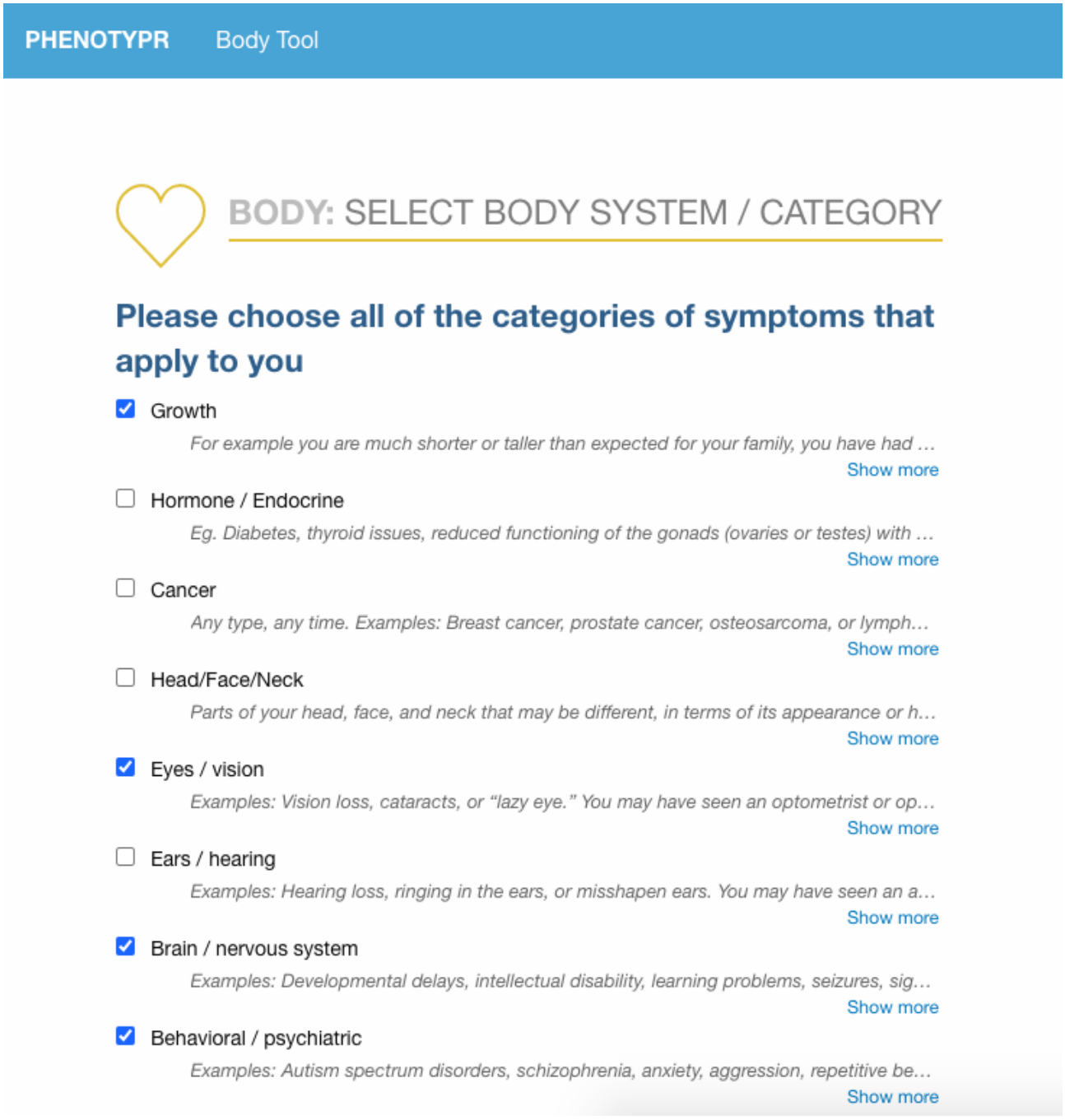


Figure 2. Data entry for Phenotypr: symptom search tool.

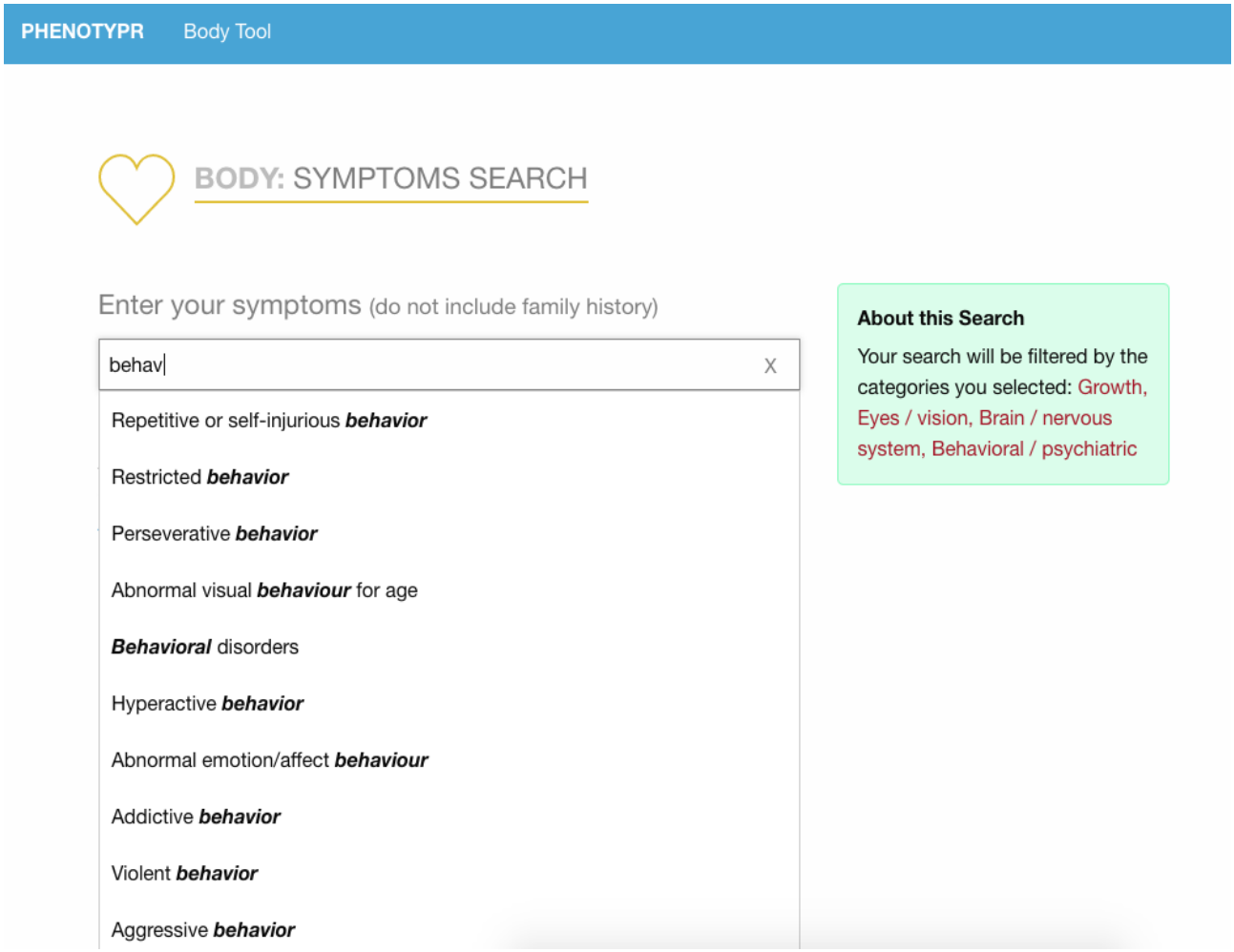
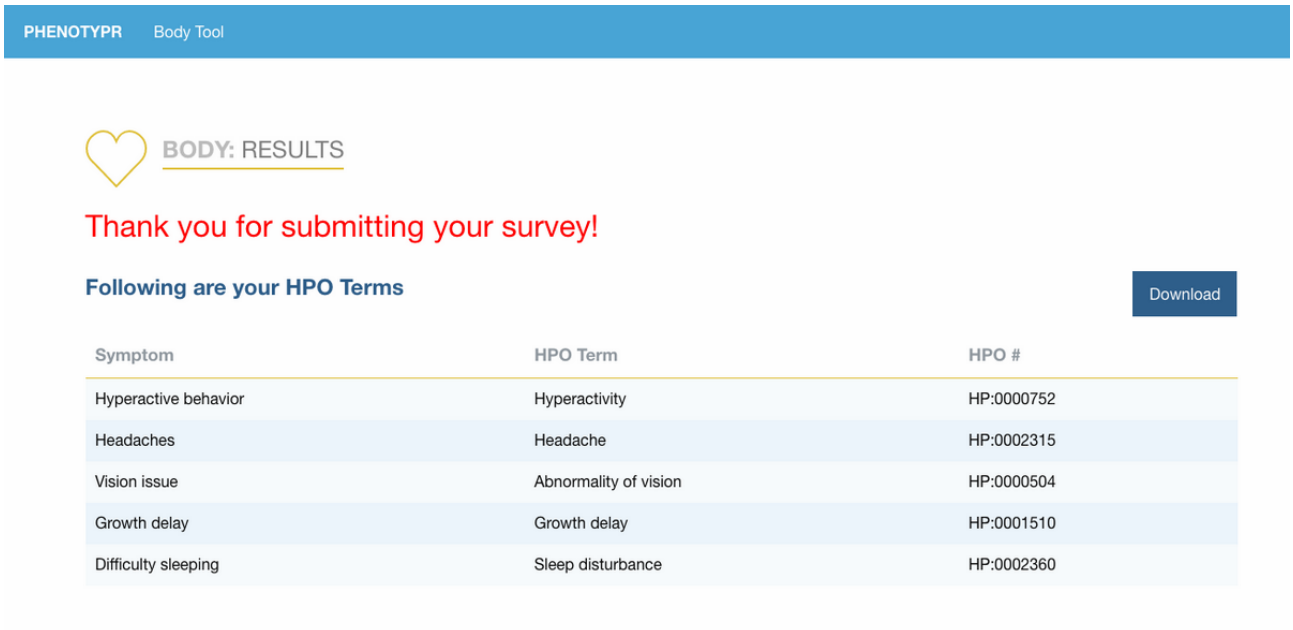


Figure 3. Example of downloadable Phenotypr output.



Data Collection

Study administrators added participants along with their unique identification number and email address into a secure REDCap database. Participants were then sent an email that contained an invitation and unique link to fill out Phenotypr. The Phenotypr survey was administered via an external web interface and took participants approximately 10 to 15 minutes to complete (Figures 1-3). Each user was assigned a unique token

that was carried through to make sure the data were tied to that user. Survey data entered into Phenotypr by participants were temporarily saved in a local Boston Children's Hospital database.

Two scripts (jobs) were run once every day. The first script imported the newly entered participant information from Boston Children's Hospital's internal instance of REDCap and synced it into the local database. The second script exported the survey data (that was filled in by the participant) from the local database into the REDCap instance. The REDCap database included all of the questions that were asked on the external web form. Thus, REDCap provided a secure way to analyze data, create participants, send reminder emails, and manage users.

Deidentified data were exported into an Excel spreadsheet for further analysis and sharing.

Results

Nineteen participants enrolled in the study and completed the survey. The caregiver, parent, or legal guardian filled out the Phenotypr survey in 15 of 19 cases. In the other four cases, it was not reported whether the respondent filled out the survey on behalf of themselves or as a caregiver, parent, or legal guardian (Table 1). Table 1 lists the phenotypic characteristics of our cohort as reported in Phenotypr (see Multimedia Appendix 1 for a complete table of phenotypic features).

In order to assess the initial accuracy of the Phenotypr tool, a comparison was conducted between the 16p13.11 microduplication syndrome phenotypes that were present in our cohort and those in a recent case report with a larger sample size [4] (Table 2). Moreover, we compared our cohort to all previously published cases of dup16p13.11 CNV and discovered several underrepresented phenotypes, which are summarized in Table 3 [1,3-5,19].

At least one of the phenotypes mentioned in previous publications, such as delayed speech, learning/intellectual disability, ASD, sleep disorder, and feeding difficulties [1-5], was reported by a majority of the Phenotypr participants (10/19 [53%] cases with at least one phenotype; Table 1). Meanwhile, aggression, which has not been widely reported in previous 16p13.11 microduplication literature, was reported by three families (Table 3).

Additionally, 4 out of 19 (21%) cases mentioned anxiety or anxiety-related behaviors, which is higher than the prevalence of anxiety reported by the Centers for Disease Control and Prevention (approximately 7% in US children aged 3-17 years) [20].

Finally, 5 out of 19 (26%) participants reported immune- and/or autoimmune-related disorders, including severe T-cell immunodeficiency in one patient and "autoimmune encephalopathy and corresponding antibody positivity" in another. Three cases reported asthma, a broadly recognized autoimmune disease, with one individual also having co-occurring hypothyroidism [21].

Table 1. Selected phenotypic features in 19 cases of 16p13.11 microduplication syndrome.

Characteristic	Presented cases (patient numbers)																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Age (years)	10	3	8	2	14	12	14	5	<1	7	6	N/A[a]	N/A	<1	18	N/A	3	9	N/A
Growth																			
Growth abnormality	–[b]	–	+[c]	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–
Muscle weakness	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–
Tall stature	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–
Development																			
Delayed speech	–	+	–	–	+	–	–	–	–	+	–	–	–	–	–	–	–	–	–
Developmental regression	–	+	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Intellectual disability	–	–	–	–	+	–	–	–	+	–	–	–	–	–	–	–	–	–	–
Mild global DD[d]	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–
Neurological and mental																			
Specific learning disability	–	–	–	–	+	–	–	+	–	–	–	–	–	–	–	–	–	–	–
Dyslexia	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–
Cognitive impairment	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–
Hypotonia	–	–	–	+	–	+	–	–	–	+	–	–	–	–	–	–	–	–	–
Poor fine motor coordination	–	–	–	–	+	–	–	–	–	+	–	–	–	–	–	–	–	–	–
Tics	–	–	–	–	–	+	+	–	–	–	–	–	–	–	–	–	–	–	–
Spasticity	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–
Dysarthria	–	–	–	+	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Anxiety	–	–	–	–	–	–	+	–	–	–	–	–	–	–	+	–	–	–	–
Depression	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–
ODD[e]	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–
ADHD[f]	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Seizure	–	–	–	–	–	–	–	+	–	+	–	–	–	–	–	–	–	–	+
ASD[g]/autistic behavior	–	–	–	–	+	+	–	–	–	+	+	–	–	–	–	–	–	–	–
Sleep disturbance	–	–	–	–	+	+	–	–	+	+	–	–	–	–	–	–	–	–	–
Parasomnia	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–
Behavior																			
Behavioral abnormality	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	+
Impulsiveness or violence	–	–	–	–	+	–	–	–	–	+	–	–	–	–	–	–	–	–	–
Aggression	–	–	–	–	+	–	–	+	–	+	–	–	–	–	–	–	–	–	–
Abnormal eating	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–
Self-mutilation	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–
Abnormal fear/anxiety	–	–	–	–	–	–	–	–	–	+	–	–	+	–	–	–	–	–	–
DMDD[h]	–	–	–	–	+	–	–	–	–	+	–	–	–	–	–	–	–	–	–
Sensory																			
Sensory impairment	–	–	–	+	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–
Hearing impairment	–	–	+	–	–	–	+	–	–	+	–	–	–	–	–	–	–	–	–
Tinnitus	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–
Astigmatism	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–

Characteristic	Presented cases (patient numbers)																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Immunity																			
Abnormality of the immune system	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Severe T-cell immunodeficiency	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Autoimmune antibody positivity	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–
Autoimmune encephalopathy	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–
Cardiac and respiratory																			
Arrhythmia	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–
Bradycardia	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Asthma	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+	–	+
Chronic lung disease	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Neonatal respiratory distress	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–
Breathing dysregulation	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–
Feeding difficulties																			
Gastrostomy tube feeding in infancy	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Dysphagia	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–
Feeding difficulties	–	–	–	+	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–

[a]N/A: not applicable.

[b]Feature absent or undisclosed.

[c]Feature present.

[d]DD: developmental delay.

[e]ODD: oppositional defiant disorder.

[f]ADHD: attention-deficit/hyperactivity disorder.

[g]ASD: autism spectrum disorder.

[h]DMDD: disruptive mood dysregulation disorder.

Table 2. Comparison of the self-phenotyping cohort with the published cohort in the study by Allach El Khattabi et al [4].

Feature	Cases in the study by Allach El Khattabi et al [4] (N=45)	Presented cases (N=19)
Hypotonia	5/45	3/19
Feeding difficulties	5/45	2/19
Neurodevelopmental features		
Developmental delay[a]	32/45	3/19
Motor delay[b]	19/45	2/19
Speech delay	35/45	3/19
Learning disabilities	30/45	2/19
ASD[c]	24/45	4/19
Aggression	—[d]	3/19
Anxiety	—	4/19
Seizures	10/45	3/19
Sleep disorders[e]	8/45	4/19
Craniofacial features		
Microcephaly	1/23	1/19
Abnormal extremities		
Hands		
Long fingers	1/22	1/19
Eyes		
Strabismus	6/45	1/19
Myopia	4/45	1/19
Amblyopia	1/23	1/19
Nystagmus	1/22	1/19
Immunological disorders		
Immunodeficiency and autoimmune diseases	—	2/19
Others		
Asthma	—	3/19
Hypothyroidism	—	1/19
Umbilical hernia	1/22	1/19
Hearing loss[f]	1/23	3/19
Abnormality of the male genitalia	2 testicular ectopia and 1 cryptorchidism	1/19

[a]Including “developmental regression” and “mild global developmental delay.”

[b]Corresponding to “poor fine motor coordination.”

[c]ASD: autism spectrum disorder.

[d]Not available.

[e]Including “sleep disturbance” and “parasomnia.”

[f]Corresponding to “hearing impairment.”

Table 3. Comparison of the self-phenotyping cohort with published case reports for anxiety, aggression, asthma, and immunological disorder phenotypes.

Phenotype	Cases in the study by Hannes et al [5] (N=5)	Cases in the study by Ramalingam et al [1] (N=8)	Cases in the study by Nagamani et al [3] (N=10)	Cases in the study by Loureiro et al [19] (N=4)	Cases in the study by Allach El Khattabi et al [4] (N=45)	Cases in this study (N=19)
Anxiety	—[a]	1/8[b]	—	1/4[c]	—	4/19
Aggression	1/5	—	1/10	—	—	3/19
Asthma	—	—	—	—	—	3/19
Immunological disorders[d]	—	—	—	—	—	2/19

[a]Not available.

[b]Co-occurring with autism spectrum disorder and attention-deficit/hyperactivity disorder.

[c]Co-occurring with autism spectrum disorder and hyperactivity.

[d]Including "abnormality of the immune system," "severe T-cell immunodeficiency," "autoimmune antibody positivity," and "autoimmune encephalopathy."

Discussion

Principal Findings

This cohort of 19 16p13.11 microduplication cases expands the knowledge of an increasingly concerning syndrome. It supports the role of many clinical features that have been previously described, including growth and behavioral disturbances, seizures, and a spectrum of neurological characteristics (Table 1). Further, the frequency of anxiety and aggression in the cohort illustrates the potential utility of self-phenotyping and nonhypothesis-driven phenotyping tools as informative data sources. The amount of immune disorders reported is also an interesting finding worthy of further investigation. The incidence rate of asthma in the cohort (3/19, 16%) was higher than that in published reports from the Centers for Disease Control and Prevention in 2018 (about 1/13, 8%) [22]. Taken together, these results imply the possibility of a broader impact of the genotype on the immune system.

Limitations

A limitation of this study is ascertainment bias. The participants were located in geographic areas where they had increased access to clinical microarray technology and formal medical diagnosis, and presented with severe enough phenotypes that the families had sought joining an online registry to participate in research; therefore, it is possible that this cohort is more severely impacted than 16p13.11 microduplication cases in the general population. Patients and families dealing with unique presentations may also be more incentivized to participate in research and use an internet self-phenotyping tool. That said, published studies may also have had similar biases, as they were recruiting participants with otherwise unexplained phenotypes, who went on to have chromosomal microarray testing, thus limiting our assessment on the presence of normal variation. Another limitation comes from the fact that for four out of 19 respondents, the individuals did not indicate whether they were filling out the survey as an individual with 16p13.11 microduplication syndrome or as a caregiver. This can add variability in the responses and cause information bias from surrogate interviews [23].

Comparison With Prior Work on 16p13.11 Microduplication

New phenotypes associated with the dup16p13.11 CNV have been continuously discovered and reported since the syndrome first came to light in 2007 when Ullmann et al initially identified the 16p13.11 microduplication and believed it predisposed patients to ASD and intellectual disability [2]. Subsequently, further studies have indicated that the 16p13.11 microduplication is likely to be involved with a wide spectrum of neurodevelopmental disorders [1,24-27]. We hereby identified two neuropsychiatric disorders, anxiety and aggression, as supplementary evidence to the underrepresented cases in prior reports, and we underline a probable high occurrence of these two phenotypes in patients with dup16p13.11. As shown in Table 3, only two cases of anxiety have been previously reported; one by Ramalingam et al (1/8 patients) [1] in 2011 and another by Loureiro et al (1/4 patients) [19] in 2017. Although one case was inherited and the other occurred *de novo*, both aforementioned cases were associated with ASD or ADHD, which aligns with literature that anxiety is a common condition in patients with ADHD and ASD [28,29]. In our cases, however, just one individual with anxiety had co-occurring ASD, whereas the remaining three cases of anxiety appeared independently. Meanwhile, aggression was also noted in three family reports, and it has gone largely unreported since Hannes et al first described a case of aggression in a patient with dup16p13.11 in 2009 [5], and another case was reported in the study by Nagamani et al in 2011 [3]. Furthermore, a potential correlation between 16p13.11 microduplication and a range of immunological disorders, specifically autoimmune conditions, was uncovered in multiple individuals in our cohort.

Aggressive behavior is one of the most common reasons for mental health referrals in children and adolescents [30-32], and can co-occur with a broad array of psychiatric and neurological illnesses, including ASD, intellectual disability, ADHD, conduct disorder, oppositional defiant disorder, disruptive mood dysregulation disorder, schizophrenia, epilepsy, anxiety, depression, and sleep disorders [33,34]. While impulsive aggression may not indicate any specific disorder, it is an important marker of severity for many psychiatric diseases [35]. For example, aggression level may affect the decision to initiate or increase medication dosage in pediatric ADHD treatment [36]. Regardless of the controversy over whether to consider

impulsive aggression as an independent categorical diagnosis like a disorder [37] or a measurable symptom secondary to some other diagnostic entity like fever or pain [32,38], it can greatly impact an individual's development and health, or even lead to high economic and medical burden for families and communities [39-41]. This paper does not aim to address the dispute over the clinical perception of aggression by joining any side or to unduly correlate 16p13.11 microduplication with the aggression phenotype. Instead, we hope to draw more attention to the repeatedly mentioned and thus important phenotypes that are self-reported by the patient families and call for more focus on related research and therapeutics.

The potential connection between neuropsychiatric disorders and autoimmunity/immunological dysfunction has received growing interest over the past decades. For instance, a nationwide population-based prospective cohort study in Denmark used a longitudinal registry to find that autoimmune diseases and infection raised the risk for subsequent mood disorder [42]. Another well-covered example is the bidirectional relationship between psychosis and autoimmune disorders [43], specifically schizophrenia and celiac disease, between which positive correlations have been suggested through studies of epidemiology, genetics, and immunology [44,45]. Notably, the discovery of autoimmune encephalitis and the disruptive autoantibody mechanisms behind it provided more direct proof of etiological linkage [46-49]. Autoimmune encephalitis is an abrupt inflammatory brain disease characterized by a variety of neuropsychiatric symptoms, such as cognitive and behavioral alterations, seizures, anxiety, and sleep disturbances [50,51], and responds to immunotherapy treatment in many cases [52,53]. Among all the possible causative autoantibodies, one of the most discussed is the N-methyl-D-aspartate receptor (NMDAR) antibody. The NMDAR antibody targets a certain subunit of NMDAR, a synaptic and neuronal cell membrane protein, and has been revealed to play a role in the development and progression of schizophrenia [47,54-57] and ASD [58-61]. Given the long-lasting belief that ASD is largely genetic [62-66], our finding of several individuals with dup16p13.11 CNV who have immunity-related diseases and neuropsychiatric symptoms, especially a case with both ASD and autoimmune encephalopathy plus corresponding antibody positivity, supports the medical field's current views and may influence future studies that seek to understand genetic origin.

It is worth noting that new associated phenotypes may emerge even for established genetic diseases as more scientific research is conducted and the demographic profiles of patient populations shift. For example, though the first case of Down syndrome was reported in the 19th century, it was not until the last two decades that individuals with Down syndrome were identified as having an increasing risk of early-onset Alzheimer disease as they age; many individuals with Down syndrome start to develop Alzheimer disease pathology in their 30s and approximately two-thirds have dementia by the age of 60 years [67,68]. Due to advancements in health care and social support, the life expectancy of the Down syndrome population has greatly improved, with the average age of death in developed countries now approaching 60 years [69]. A similar increase in knowledge has occurred regarding Turner syndrome, which was first described in the 1930s and is one of the most common genetic disorders [70,71]. Some patients with Turner syndrome who carried a mosaic 46, XY karyotype or an abnormal Y chromosome were recently found to have a higher risk of developing gonadoblastoma and other gonadal tumors owing to the widespread use and easy availability of polymerase chain reaction technology [7,72]. For genetic diseases with variable presentations and complex genotypes, such as Down syndrome and Turner syndrome, it may be too early to announce that all phenotypes have been exhaustively discovered and included.

We do not claim that Phenotypr is a substitute for clinician phenotyping, and it is unlikely that participants will be able to describe some of the highly technical aspects of their disorders in clinical terminology. Therefore, we do not consider the underreporting of some phenotypes in Phenotypr to be notable. In contrast, phenotypes that were repeatedly mentioned by participants but have not been well discussed in the literature are worthy of further investigation. It is possible that patients and families are not mentioning these phenotypes in clinic visits or that their existing concerns are not being fully understood.

Conclusions

In this study, we utilized the Phenotypr tool to collect self-phenotyping data from 19 16p13.11 microduplication syndrome cases, with the aim of identifying underrepresented phenotypes in the current scientific literature. A number of phenotypes were highlighted, including aggression, anxiety, and a range of immunological disorders. In addition to the typically recognized phenotypes, dup16p13.11 CNV showed a stronger predisposition to aggression and anxiety compared to previously reported cases, as both phenotypes were mentioned by multiple Phenotypr participant families (4/19, 21%). Moreover, we found that three out of four cases with anxiety did not have co-occurring ASD or ADHD, which differed from two other published cases [1,28]. An interesting case that involved ASD and autoimmune encephalopathy with corresponding antibody positivity was also identified using Phenotypr. These findings illustrate some important hypotheses. First, aggression and anxiety may be more common than previously understood in 16p13.11 microduplication cases. Second, anxiety may appear independently as a result of dup16p13.11 CNV instead of being accompanied with ASD or ADHD. Third, immune and autoimmune disorders might be phenotypes of 16p13.11 microduplication, and dup16p13.11 CNV might play a genetic role in the association between autoimmune encephalopathy and ASD. Fourth, in consideration of the incomplete penetrance and varied expression of this syndrome in a broad spectrum of neuropsychiatric disorders, patients and their clinicians should be aware of all possible phenotypes to ensure that treatment is as effective as possible. Fifth, HPO and layperson HPO profiles acquired through patient self-phenotyping can serve as a valuable data source for the exploration of underreported phenotypes in the scientific literature, especially for rare disorders with variable presentations.

Future work will apply Phenotypr to additional 16p13.11 microduplication cases and correlate phenotypic results to the size of the duplicated interval and the genes involved. We also

look forward to adopting Phenotypr as a complementary data source in other genetic cohort studies of rare diseases.

Acknowledgments

We thank the patients and families who participated in this study. This work was supported by the Patient-Centered Outcomes Research Institute (PCORI) (grant #HSRP20181624), Tommy Fuss Fund, Robin and Jonathan Klein Family, and National Institutes of Health U54 (HD090255).

Conflicts of Interest

In the past 3 years, JGH has received grant support from the Tommy Fuss Fund. He is the founding head of the Scientific Advisory Board and has equity in Mightier/Neuro'motion, Inc, a company working on emotional regulation training tools. He has been a consultant to Neurocrine and Alkermes pharmaceutical companies. Author MH is the cofounder of Pryzm Health. The other authors have no conflicts to declare.

Multimedia Appendix 1
Selected phenotypic features in 19 cases of 16p13.11 microduplication syndrome.
[DOCX File , 56 KB - jmir_v23i3e21023_app1.docx]

References

1. Ramalingam A, Zhou X, Fiedler SD, Brawner SJ, Joyce JM, Liu H, et al. 16p13.11 duplication is a risk factor for a wide spectrum of neuropsychiatric disorders. J Hum Genet 2011 Jul;56(7):541-544 [FREE Full text] [doi: 10.1038/jhg.2011.42] [Medline: 21614007]
2. Ullmann R, Turner G, Kirchhoff M, Chen W, Tonge B, Rosenberg C, et al. Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. Hum Mutat 2007 Jul;28(7):674-682. [doi: 10.1002/humu.20546] [Medline: 17480035]
3. Nagamani SCS, Erez A, Bader P, Lalani SR, Scott DA, Scaglia F, et al. Phenotypic manifestations of copy number variation in chromosome 16p13.11. Eur J Hum Genet 2011 Mar;19(3):280-286 [FREE Full text] [doi: 10.1038/ejhg.2010.184] [Medline: 21150890]
4. Allach El Khattabi L, Heide S, Caberg J, Andrieux J, Doco Fenzy M, Vincent-Delorme C, et al. 16p13.11 microduplication in 45 new patients: refined clinical significance and genotype-phenotype correlations. J Med Genet 2020 May;57(5):301-307. [doi: 10.1136/jmedgenet-2018-105389] [Medline: 30287593]
5. Hannes FD, Sharp AJ, Mefford HC, de Ravel T, Ruivenkamp CA, Breuning MH, et al. Recurrent reciprocal deletions and duplications of 16p13.11: the deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant. J Med Genet 2009 Apr;46(4):223-232 [FREE Full text] [doi: 10.1136/jmg.2007.055202] [Medline: 18550696]
6. Brownstein CA, Kleiman RJ, Engle EC, Towne MC, D'Angelo EJ, Yu TW, et al. Overlapping 16p13.11 deletion and gain of copies variations associated with childhood onset psychosis include genes with mechanistic implications for autism associated pathways: Two case reports. Am J Med Genet A 2016 May;170A(5):1165-1173 [FREE Full text] [doi: 10.1002/ajmg.a.37595] [Medline: 26887912]
7. Tropeano M, Andrieux J, Collier DA. Clinical utility gene card for: 16p13.11 microdeletion syndrome. Eur J Hum Genet 2014 May;22(5):1-4 [FREE Full text] [doi: 10.1038/ejhg.2013.230] [Medline: 24105370]
8. Bradshaw NJ, Ukkola-Vuoti L, Pankakoski M, Zheutlin AB, Ortega-Alonso A, Torniainen-Holm M, et al. The genomic locus can affect treatment of psychiatric illness through gene expression changes related to microRNA-484. Open Biol 2017 Nov;7(11) [FREE Full text] [doi: 10.1098/rsob.170153] [Medline: 29142105]
9. Savatt JM, Azzariti DR, Faucett WA, Harrison S, Hart J, Kattman B, et al. ClinGen's GenomeConnect registry enables patient-centered data sharing. Hum Mutat 2018 Nov;39(11):1668-1676 [FREE Full text] [doi: 10.1002/humu.23633] [Medline: 30311371]
10. Kirkpatrick BE, Riggs ER, Azzariti DR, Miller VR, Ledbetter DH, Miller DT, ClinGen Resource. GenomeConnect: matchmaking between patients, clinical laboratories, and researchers to improve genomic knowledge. Hum Mutat 2015 Oct;36(10):974-978 [FREE Full text] [doi: 10.1002/humu.22838] [Medline: 26178529]
11. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gourdine J, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res 2019 Jan 08;47(D1):D1018-D1027 [FREE Full text] [doi: 10.1093/nar/gky1105] [Medline: 30476213]
12. Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Bailleul-Forestier I, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res 2014 Jan;42(Database issue):D966-D974 [FREE Full text] [doi: 10.1093/nar/gkt1026] [Medline: 24217912]

13. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet 2008 Nov;83(5):610-615 [FREE Full text] [doi: 10.1016/j.ajhg.2008.09.017] [Medline: 18950739]
14. Vasilevsky NA, Foster ED, Engelstad ME, Carmody L, Might M, Chambers C, et al. Plain-language medical vocabulary for precision diagnosis. Nat Genet 2018 Apr;50(4):474-476 [FREE Full text] [doi: 10.1038/s41588-018-0096-x] [Medline: 29632381]
15. Phenotypr. Boston Children's Hospital. 2017. URL: http://www.phenotypr.com [accessed 2020-07-28]
16. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
17. Solr. The Apache Software Foundation. URL: https://lucene.apache.org/solr/ [accessed 2020-07-28]
18. Vue.js. 2020. URL: https://vuejs.org/ [accessed 2020-07-28]
19. Loureiro S, Almeida J, Café C, Conceição I, Mouga S, Beleza A, et al. Copy number variations in chromosome 16p13.11-The neurodevelopmental clinical spectrum. Current Pediatric Research 2017 Jan;21(1):116-129 [FREE Full text]
20. Data and Statistics on Children's Mental Health. Centers for Disease Control and Prevention. 2019 Jun 15. URL: https://www.cdc.gov/childrensmentalhealth/data.html [accessed 2020-07-28]
21. Amino N. 4 Autoimmunity and hypothyroidism. Baillière's Clinical Endocrinology and Metabolism 1988 Aug;2(3):591-617. [doi: 10.1016/s0950-351x(88)80055-7] [Medline: 3066320]
22. Most Recent National Asthma Data. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm [accessed 2020-08-31]
23. Aitken JF, Green A, MacLennan R, Jackman L, Martin NG. Comparability of surrogate and self-reported information on melanoma risk factors. Br J Cancer 1993 May;67(5):1036-1041 [FREE Full text] [doi: 10.1038/bjc.1993.190] [Medline: 8494696]
24. Mefford HC, Cooper GM, Zerr T, Smith JD, Baker C, Shafer N, et al. A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. Genome Res 2009 Sep;19(9):1579-1585 [FREE Full text] [doi: 10.1101/gr.094987.109] [Medline: 19506092]
25. Mefford HC, Muhle H, Ostertag P, von Spiczak S, Buysse K, Baker C, et al. Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. PLoS Genet 2010 May 20;6(5):e1000962 [FREE Full text] [doi: 10.1371/journal.pgen.1000962] [Medline: 20502679]
26. Williams NM, Zaharieva I, Martin A, Langley K, Mantripragada K, Fossdal R, et al. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. The Lancet 2010 Oct 23;376(9750):1401-1408 [FREE Full text] [doi: 10.1016/S0140-6736(10)61109-9] [Medline: 20888040]
27. Kuang S, Guo D, Prakash SK, McDonald MN, Johnson RJ, Wang M, GenTAC Investigators. Recurrent chromosome 16p13.1 duplications are a risk factor for aortic dissections. PLoS Genet 2011 Jun;7(6):e1002118 [FREE Full text] [doi: 10.1371/journal.pgen.1002118] [Medline: 21698135]
28. Schatz DB, Rostain AL. ADHD with comorbid anxiety: a review of the current literature. J Atten Disord 2006 Nov;10(2):141-149. [doi: 10.1177/1087054706286698] [Medline: 17085624]
29. South M, Rodgers J, Van Hecke A. Anxiety and ASD: Current Progress and Ongoing Challenges. J Autism Dev Disord 2017 Dec;47(12):3679-3681. [doi: 10.1007/s10803-017-3322-y] [Medline: 28988366]
30. Connor DF. Aggression and Antisocial Behavior in Children and Adolescents: Research and Treatment. New York, NY: Guilford Press; 2004.
31. Steiner H, Saxena K, Chang K. Psychopharmacologic strategies for the treatment of aggression in juveniles. CNS Spectr 2003 Apr;8(4):298-308. [doi: 10.1017/s1092852900018514] [Medline: 12679744]
32. Jensen PS, Youngstrom EA, Steiner H, Findling RL, Meyer RE, Malone RP, et al. Consensus report on impulsive aggression as a symptom across diagnostic categories in child psychiatry: implications for medication studies. J Am Acad Child Adolesc Psychiatry 2007 Mar;46(3):309-322. [doi: 10.1097/chi.0b013e31802f1454] [Medline: 17314717]
33. Connor DF, Newcorn JH, Saylor KE, Amann BH, Scahill L, Robb AS, et al. Maladaptive Aggression: With a Focus on Impulsive Aggression in Children and Adolescents. J Child Adolesc Psychopharmacol 2019 Oct;29(8):576-591 [FREE Full text] [doi: 10.1089/cap.2019.0039] [Medline: 31453715]
34. Magalotti S, Neudecker M, Zaraa S, McVoy M. Understanding Chronic Aggression and Its Treatment in Children and Adolescents. Curr Psychiatry Rep 2019 Nov 18;21(12):123 [FREE Full text] [doi: 10.1007/s11920-019-1105-1] [Medline: 31741142]
35. Connor DF, McLaughlin TJ. Aggression and diagnosis in psychiatrically referred children. Child Psychiatry Hum Dev 2006;37(1):1-14. [doi: 10.1007/s10578-006-0015-8] [Medline: 16779500]
36. Connor DF, Doerfler LA. ADHD with comorbid oppositional defiant disorder or conduct disorder: discrete or nondistinct disruptive behavior disorders? J Atten Disord 2008 Sep;12(2):126-134. [doi: 10.1177/1087054707308486] [Medline: 17934178]
37. Coccaro EF, Fanning JR, Phan KL, Lee R. Serotonin and impulsive aggression. CNS Spectr 2015 Jun;20(3):295-302. [doi: 10.1017/S1092852915000310] [Medline: 25997605]

38. Raine A, Dodge K, Loeber R, Gatzke-Kopp L, Lynam D, Reynolds C, et al. The Reactive-Proactive Aggression Questionnaire: Differential Correlates of Reactive and Proactive Aggression in Adolescent Boys. Aggress Behav 2006 Apr 01;32(2):159-171 [FREE Full text] [doi: 10.1002/ab.20115] [Medline: 20798781]
39. Broidy LM, Nagin DS, Tremblay RE, Bates JE, Brame B, Dodge KA, et al. Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. Developmental Psychology 2003 Mar;39(2):222-245 [FREE Full text] [doi: 10.1037//0012-1649.39.2.222] [Medline: 12661883]
40. Raaijmakers MA, Posthumus JA, van Hout BA, van Engeland H, Matthys W. Cross-sectional study into the costs and impact on family functioning of 4-year-old children with aggressive behavior. Prev Sci 2011 Jun;12(2):192-200 [FREE Full text] [doi: 10.1007/s11121-011-0204-y] [Medline: 21369841]
41. Hendricks K, Liu J. Childbearing depression and childhood aggression: literature review. MCN Am J Matern Child Nurs 2012;37(4):253-261 [FREE Full text] [doi: 10.1097/NMC.0b013e3182482c5e] [Medline: 22739482]
42. Benros ME, Waltoft BL, Nordentoft M, Ostergaard SD, Eaton WW, Krogh J, et al. Autoimmune diseases and severe infections as risk factors for mood disorders: a nationwide study. JAMA Psychiatry 2013 Aug;70(8):812-820. [doi: 10.1001/jamapsychiatry.2013.1111] [Medline: 23760347]
43. Benros ME, Nielsen PR, Nordentoft M, Eaton WW, Dalton SO, Mortensen PB. Autoimmune diseases and severe infections as risk factors for schizophrenia: a 30-year population-based register study. Am J Psychiatry 2011 Dec;168(12):1303-1310. [doi: 10.1176/appi.ajp.2011.11030516] [Medline: 22193673]
44. Benros ME, Eaton WW, Mortensen PB. The epidemiologic evidence linking autoimmune diseases and psychosis. Biol Psychiatry 2014 Feb 15;75(4):300-306. [doi: 10.1016/j.biopsych.2013.09.023] [Medline: 24199668]
45. Jeppesen R, Benros ME. Autoimmune Diseases and Psychotic Disorders. Front Psychiatry 2019 Mar 20;10:131 [FREE Full text] [doi: 10.3389/fpsyt.2019.00131] [Medline: 30949074]
46. Dalmau J, Tüzün E, Wu H, Masjuan J, Rossi JE, Voloschin A, et al. Paraneoplastic anti-N-methyl-D-aspartate receptor encephalitis associated with ovarian teratoma. Ann Neurol 2007 Jan;61(1):25-36 [FREE Full text] [doi: 10.1002/ana.21050] [Medline: 17262855]
47. Kayser MS, Dalmau J. The emerging link between autoimmune disorders and neuropsychiatric disease. J Neuropsychiatry Clin Neurosci 2011;23(1):90-97 [FREE Full text] [doi: 10.1176/jnp.23.1.jnp90] [Medline: 21304144]
48. Al-Diwani AAJ, Pollak TA, Irani SR, Lennox BR. Psychosis: an autoimmune disease? Immunology 2017 Nov;152(3):388-401 [FREE Full text] [doi: 10.1111/imm.12795] [Medline: 28704576]
49. Mané-Damas M, Hoffmann C, Zong S, Tan A, Molenaar PC, Losen M, et al. Autoimmunity in psychotic disorders. Where we stand, challenges and opportunities. Autoimmun Rev 2019 Sep;18(9):102348 [FREE Full text] [doi: 10.1016/j.autrev.2019.102348] [Medline: 31323365]
50. Dalmau J, Gleichman AJ, Hughes EG, Rossi JE, Peng X, Lai M, et al. Anti-NMDA-receptor encephalitis: case series and analysis of the effects of antibodies. The Lancet Neurology 2008 Dec;7(12):1091-1098 [FREE Full text] [doi: 10.1016/S1474-4422(08)70224-2] [Medline: 18851928]
51. Herken J, Prüss H. Red Flags: Clinical Signs for Identifying Autoimmune Encephalitis in Psychiatric Patients. Front Psychiatry 2017;8:25 [FREE Full text] [doi: 10.3389/fpsyt.2017.00025] [Medline: 28261116]
52. Gotkine M, Ben-Hur T, Vincent A, Vaknin-Dembinsky A. Limbic encephalitis presenting as a post-partum psychiatric condition. J Neurol Sci 2011 Sep 15;308(1-2):152-154. [doi: 10.1016/j.jns.2011.06.017] [Medline: 21705029]
53. Kayser MS, Titulaer MJ, Gresa-Arribas N, Dalmau J. Frequency and characteristics of isolated psychiatric episodes in anti–N-methyl-d-aspartate receptor encephalitis. JAMA Neurol 2013 Sep 01;70(9):1133-1139 [FREE Full text] [doi: 10.1001/jamaneurol.2013.3216] [Medline: 23877059]
54. Zandi MS, Irani SR, Lang B, Waters P, Jones PB, McKenna P, et al. Disease-relevant autoantibodies in first episode schizophrenia. J Neurol 2011 Apr;258(4):686-688 [FREE Full text] [doi: 10.1007/s00415-010-5788-9] [Medline: 20972895]
55. Steiner J, Walter M, Glanz W, Sarnyai Z, Bernstein H, Vielhaber S, et al. Increased prevalence of diverse N-methyl-D-aspartate glutamate receptor antibodies in patients with an initial diagnosis of schizophrenia: specific relevance of IgG NR1a antibodies for distinction from N-methyl-D-aspartate glutamate receptor encephalitis. JAMA Psychiatry 2013 Mar;70(3):271-278. [doi: 10.1001/2013.jamapsychiatry.86] [Medline: 23344076]
56. Pearlman DM, Najjar S. Meta-analysis of the association between N-methyl-d-aspartate receptor antibodies and schizophrenia, schizoaffective disorder, bipolar disorder, and major depressive disorder. Schizophr Res 2014 Aug;157(1-3):249-258 [FREE Full text] [doi: 10.1016/j.schres.2014.05.001] [Medline: 24882425]
57. Lennox BR, Palmer-Cooper EC, Pollak T, Hainsworth J, Marks J, Jacobson L, et al. Prevalence and clinical characteristics of serum neuronal cell surface antibodies in first-episode psychosis: a case-control study. The Lancet Psychiatry 2017 Jan;4(1):42-48 [FREE Full text] [doi: 10.1016/S2215-0366(16)30375-3] [Medline: 27965002]
58. Tarabeux J, Kebir O, Gauthier J, Hamdan FF, Xiong L, Piton A, S2D team, et al. Rare mutations in N-methyl-D-aspartate glutamate receptors in autism spectrum disorders and schizophrenia. Transl Psychiatry 2011 Nov 15;1:e55 [FREE Full text] [doi: 10.1038/tp.2011.52] [Medline: 22833210]
59. Creten C, van der Zwaan S, Blankespoor RJ, Maatkamp A, Nicolai J, van Os J, et al. Late onset autism and anti-NMDA-receptor encephalitis. The Lancet 2011 Jul 02;378(9785):98. [doi: 10.1016/S0140-6736(11)60548-5] [Medline: 21724038]

60. Hacohen Y, Wright S, Gadian J, Vincent A, Lim M, Wassmer E, et al. N-methyl-d-aspartate (NMDA) receptor antibodies encephalitis mimicking an autistic regression. Dev Med Child Neurol 2016 Oct 03;58(10):1092-1094 [FREE Full text] [doi: 10.1111/dmcn.13169] [Medline: 27255282]
61. Tzang R, Chang C, Chang Y, Lane H. Autism Associated With Anti-NMDAR Encephalitis: Glutamate-Related Therapy. Front Psychiatry 2019 Jun 21;10:440 [FREE Full text] [doi: 10.3389/fpsyt.2019.00440] [Medline: 31293459]
62. Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, et al. Autism as a strongly genetic disorder: evidence from a British twin study. Psychol Med 1995 Jan 09;25(1):63-77. [doi: 10.1017/s0033291700028099] [Medline: 7792363]
63. Comi AM, Zimmerman AW, Frye VH, Law PA, Peeden JN. Familial clustering of autoimmune disorders and evaluation of medical risk factors in autism. J Child Neurol 1999 Jun 02;14(6):388-394. [doi: 10.1177/088307389901400608] [Medline: 10385847]
64. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong association of de novo copy number mutations with autism. Science 2007 Apr 20;316(5823):445-449 [FREE Full text] [doi: 10.1126/science.1138659] [Medline: 17363630]
65. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature 2009 May 28;459(7246):528-533 [FREE Full text] [doi: 10.1038/nature07999] [Medline: 19404256]
66. Ronald A, Hoekstra RA. Autism spectrum disorders and autistic traits: a decade of new twin studies. Am J Med Genet B Neuropsychiatr Genet 2011 Apr 13;156B(3):255-274. [doi: 10.1002/ajmg.b.31159] [Medline: 21438136]
67. McCarron M, McCallion P, Reilly E, Mulryan N. A prospective 14-year longitudinal follow-up of dementia in persons with Down syndrome. J Intellect Disabil Res 2014 Jan;58(1):61-70. [doi: 10.1111/jir.12074] [Medline: 23902161]
68. Head E, Powell D, Gold BT, Schmitt FA. Alzheimer's Disease in Down Syndrome. Eur J Neurodegener Dis 2012 Dec;1(3):353-364 [FREE Full text] [Medline: 25285303]
69. Glasson EJ, Sullivan SG, Hussain R, Petterson BA, Montgomery PD, Bittles AH. The changing survival profile of people with Down's syndrome: implications for genetic counselling. Clin Genet 2002 Nov;62(5):390-393. [doi: 10.1034/j.1399-0004.2002.620506.x] [Medline: 12431254]
70. Turner HH. A syndrome of infantilism, congenital webbed neck, and cubitus valgus. Endocrinology 1938 Nov;23(5):566-574. [doi: 10.1210/endo-23-5-566]
71. Stochholm K, Juul S, Juel K, Naeraa RW, Gravholt CH. Prevalence, incidence, diagnostic delay, and mortality in Turner syndrome. J Clin Endocrinol Metab 2006 Oct;91(10):3897-3902. [doi: 10.1210/jc.2006-0558] [Medline: 16849410]
72. Cools M, Pleskacova J, Stoop H, Hoebeke P, Van Laecke E, Drop SLS, Mosaicism Collaborative Group. Gonadal pathology and tumor risk in relation to clinical characteristics in patients with 45,X/46,XY mosaicism. J Clin Endocrinol Metab 2011 Jul;96(7):E1171-E1180. [doi: 10.1210/jc.2011-0232] [Medline: 21508138]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ASD: autism spectrum disorder
CNV: copy number variation
HPO: human phenotype ontology
NMDAR: N-methyl-D-aspartate receptor
REDCap: Research Electronic Data Capture

Edited by G Eysenbach; submitted 13.10.20; peer-reviewed by AJ Nagarajan; comments to author 20.11.20; revised version received 26.12.20; accepted 16.01.21; published 16.03.21.

Please cite as:
Li J, Hojlo MA, Chennuri S, Gujral N, Paterson HL, Shefchek KA, Genetti CA, Cohn EL, Sewalk KC, Garvey EA, Buttermore ED, Anderson NC, Beggs AH, Agrawal PB, Brownstein JS, Haendel MA, Holm IA, Gonzalez-Heydrich J, Brownstein CA
Underrepresentation of Phenotypic Variability of 16p13.11 Microduplication Syndrome Assessed With an Online Self-Phenotyping Tool (Phenotypr): Cohort Study
J Med Internet Res 2021;23(3):e21023
URL: https://www.jmir.org/2021/3/e21023
doi:10.2196/21023
PMID:33724192

Letter to the Editor

Practical Considerations and Successful Implementation of Vital Signs Monitoring. Comment on "Continuous Versus Intermittent Vital Signs Monitoring Using a Wearable, Wireless Patch in Patients Admitted to Surgical Wards: Pilot Cluster Randomized Controlled Trial"

Caoimhe Walsh[1]; David Zargaran[2]; Nikhil Patel[3]; Amelia White[4]; Foteini Stefania Koumpa[5]; Ravina Tanna[6]; Muhammad Arsalan Ashraf[2]

[1]Epsom and St Helier University Hospitals NHS Trust, Surrey, United Kingdom
[2]Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom
[3]Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom
[4]Royal Free London NHS Foundation Trust, London, United Kingdom
[5]London North West University Healthcare NHS Trust, London, United Kingdom
[6]Luton and Dunstable University Hospital NHS Foundation Trust, London, United Kingdom

Corresponding Author:
Caoimhe Walsh
Epsom and St Helier University Hospitals NHS Trust
Wrythe Lane, Carshalton
Surrey, SM5 1AA
United Kingdom
Phone: 44 7584061131
Email: caoimhemwalsh@gmail.com

Related Article:

Comment on: https://www.jmir.org/2018/12/e10802/

(J Med Internet Res 2021;23(3):e14042) doi:10.2196/14042

KEYWORDS

general surgery; monitoring; observations; vital signs

We read with great interest the study by Downey et al [1], which examined the clinical utility and practicality of introducing a wearable, wireless patch device for continuous vital signs monitoring in surgical inpatients. As surgical trainees, we frequently rely on the recording of patients' vital signs on observation charts and the use of early warning scores. These systems are in place as a safety mechanism to highlight acutely unwell patients and those at risk of clinical deterioration [2]. However, we are also all too familiar with the limitations posed by current methods of collecting and communicating this critical information, and there is a real need for further improvements in both of these areas [3].

Having a continuous monitoring device poses a number of potential benefits. First, by having many more data points clinicians can more accurately define long-term trends for each patient. Even more importantly, deterioration can be highlighted much sooner than currently permitted due to the 5-6–hour intervals between "observation rounds"—but will this result in information overload? It is easy to capture information, but this can be misleading, and overinterpretation of incidental or minor alterations in vital signs may lead to unnecessary additional investigations for a patient. Furthermore, although data are being collected continuously, nursing staff can only check this at discrete time points due to other jobs and ward responsibilities. Therefore, staff may nevertheless fail to recognize unwell patients at the earliest time point. This is evidenced by the fact that it still took an average of 626 minutes to initiate treatment for sepsis in the patient cohort with continuous monitoring, despite the UK national guidelines for the management of sepsis stating that antibiotics should be administered within 1 hour of the patient first being suspected to have sepsis [4].

We understand that the software initially overwhelmed nursing staff with false alerts, which highlights the important yet challenging balance to make between false alerts and detecting

significant clinical changes. We would be interested to understand what parameters were used for the vital signs alerts, and whether these parameters were standardized across all patients or relative to the individual patient's baseline vital signs. Furthermore, does the alert system indicate the severity of the trigger?

There is also the consideration of how practical it is for patients to wear such devices continuously around their chest. Anecdotally, we know that patients who are advised to wear TED (thromboembolism-deterrent) stockings as part of venous thromboembolism prophylaxis are often found to have removed them "temporarily" due to discomfort or to have a shower, with subsequent difficulty putting them back on. We can envisage similar factors affecting the wearing of these devices for vital signs monitoring; for example, they may be removed if they interfere with clinical examination of the chest, or again due to comfort or personal hygiene reasons. We already anticipate this happening as the study noted that 24% of patients did not wear the wireless patch for the whole length of their admission. This means that as well as providing training for nurses, doctors and indeed patients themselves would also need to be trained in how to position the devices on the chest to ensure they are placed back correctly if removed for whatever reason.

One solution to minimize the impact of wearing such devices would be to reserve them for use out of hours. Typically, this is when wards are minimally staffed and clinicians are individually responsible for a far larger number of patients, leading to a more significant need to highlight deteriorating patients [5]. During this time, patients are typically less active therefore there would be fewer false alarms, for example, from increased heart rate due to patients mobilizing, and with patients unlikely to be showering overnight, which would result in reduced chances for the need to remove or replace the device. Within normal working hours, nurses and doctors have a much greater presence in the ward; therefore, concerns regarding unwell patients are more likely to be raised in a timely manner. Thus, it may be that during the day, current methods of intermittently collecting vital signs data will continue to suffice; this would also reduce the interference of wearing the device with activities such as showering or being clinically examined.

We recommend that future studies control for the time of day when the clinical deterioration occurred. By analyzing separately the time to treat sepsis both in hours and out of hours for the continuously monitored and intermittently monitored groups, it would be possible to identify more objectively whether there is a particular time of day where continuous vital signs monitoring renders the greatest clinical benefit over intermittent monitoring.

This paper [1] has identified a sensible and considered solution to the issue of collecting and communicating vital signs data on surgical inpatients. Despite further work being required to streamline the implementation of this system into clinical use, we commend the authors on their innovative device. We hope this technology will soon help to improve clinical outcomes and look forward to seeing a study with a more significant population size, and thus greater power, to enable stronger conclusions to be drawn from the results.

Editorial Notice

The corresponding author of "Continuous Versus Intermittent Vital Signs Monitoring Using a Wearable, Wireless Patch in Patients Admitted to Surgical Wards: Pilot Cluster Randomized Controlled Trial" did not indicate a desire to respond.

Conflicts of Interest

None declared.

References

1. Downey C, Randell R, Brown J, Jayne DG. Continuous Versus Intermittent Vital Signs Monitoring Using a Wearable, Wireless Patch in Patients Admitted to Surgical Wards: Pilot Cluster Randomized Controlled Trial. J Med Internet Res 2018 Dec 11;20(12):e10802 [FREE Full text] [doi: 10.2196/10802] [Medline: 30538086]
2. National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS. Updated report of a working party. Royal College of Physicians. London, UK: RCP; 2017. URL: https://www.rcplondon.ac.uk/file/8504/download?token=HBEWW78k [accessed 2021-02-10]
3. Downey CL, Tahir W, Randell R, Brown JM, Jayne DG. Strengths and limitations of early warning scores: A systematic review and narrative synthesis. Int J Nurs Stud 2017 Nov;76:106-119. [doi: 10.1016/j.ijnurstu.2017.09.003] [Medline: 28950188]
4. Sepsis Quality Standard 2017. National Institute for Health and Care Excellence. 2017. URL: https://www.nice.org.uk/guidance/qs161/resources/sepsis-pdf-75545595402181 [accessed 2021-02-10]
5. Hands C, Reid E, Meredith P, Smith GB, Prytherch DR, Schmidt PE, et al. Patterns in the recording of vital signs and early warning scores: compliance with a clinical escalation protocol. BMJ Qual Saf 2013 Sep;22(9):719-726. [doi: 10.1136/bmjqs-2013-001954] [Medline: 23603474]

Abbreviations

TED: thromboembolism-deterrent

Edited by G Eysenbach, T Derrick; submitted 17.03.19; peer-reviewed by C Downey; accepted 27.01.21; published 11.03.21.

Please cite as:
Walsh C, Zargaran D, Patel N, White A, Koumpa FS, Tanna R, Ashraf MA
Practical Considerations and Successful Implementation of Vital Signs Monitoring. Comment on "Continuous Versus Intermittent Vital Signs Monitoring Using a Wearable, Wireless Patch in Patients Admitted to Surgical Wards: Pilot Cluster Randomized Controlled Trial"
J Med Internet Res 2021;23(3):e14042
URL: https://www.jmir.org/2021/3/e14042
doi:10.2196/14042
PMID:33704079

Corrigenda and Addenda

Correction: Using the Patient Portal Sexual Health Instrument in Surveys and Patient Questionnaires Among Sexual Minority Men in the United States: Cross-sectional Psychometric Validation Study

Kevon-Mark P Jackman[1], DrPH, MPH; Jeremy Kane[2], PhD, MPH; Hadi Kharrazi[3], PhD, MD; Renee M Johnson[1], PhD, MPH; Carl Latkin[4], MS, PhD

[1]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Department of Epidemiology, Columbia University Mailman School of Public Health, NY, NY, United States

[3]Department of Health Policy and Management, Center for Population Health IT, Johns Hopkins Bloomberg School of Public Health, Batimore, MD, United States

[4]Department of Health, Behavior, and Society, Johns Hopkins Blooomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Kevon-Mark P Jackman, DrPH, MPH
Department of Mental Health
Johns Hopkins Bloomberg School of Public Health
624 N Broadway
Hampton House 8th Floor
Baltimore, MD, 21205
United States
Phone: 1 410 955 3910
Email: kjackma2@jhmi.edu

Related Article:

Correction of: https://www.jmir.org/2021/2/e18750/

(J Med Internet Res 2021;23(3):e28358) doi:10.2196/28358

In "Using the Patient Portal Sexual Health Instrument in Surveys and Patient Questionnaires Among Sexual Minority Men in the United States: Cross-sectional Psychometric Validation Study" (J Med Internet Res 2021;23(2):e18750) four errors were noted.

In the originally published paper, the "greater than or equal to" symbol (≥) was missing in three places in tables and in one place in the main text due to an XML conversion error. The following corrections have been made:

In Table 1, under "Age (years)," "40" has been corrected to "≥40".

In Table 5, under "Age category," "40" has been corrected to "≥40".

In Table 6, under "Age category," "40" has been corrected to "≥40".

In the section "AMIS-PPSHI Scores," the symbol was missing in the following sentence:

> *However, scores were marginally higher among participants with a Kessler 6-item psychological distress scale (K6) score13.*

This has been corrected to:

> *However, scores were marginally higher among participants with a Kessler 6-item psychological distress scale (K6) score≥13.*

The correction will appear in the online version of the paper on the JMIR Publications website on March 5, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 02.03.21; this is a non–peer-reviewed article;accepted 03.03.21; published 05.03.21.

Please cite as:
Jackman KMP, Kane J, Kharrazi H, Johnson RM, Latkin C
Correction: Using the Patient Portal Sexual Health Instrument in Surveys and Patient Questionnaires Among Sexual Minority Men in the United States: Cross-sectional Psychometric Validation Study
J Med Internet Res 2021;23(3):e28358
URL: https://www.jmir.org/2021/3/e28358
doi:10.2196/28358
PMID:33667179

Corrigenda and Addenda

Correction: Association Between Institutional Social Media Involvement and Gastroenterology Divisional Rankings: Cohort Study

Austin Lee Chiang[1,2], MD, MPH; Loren Galler Rabinowitz[3], MD; Akhil Kumar[4], BS; Walter Wai-Yip Chan[1,2], MD, MPH

[1]Brigham and Women's Hospital, Boston, MA, United States
[2]Harvard Medical School, Boston, MA, United States
[3]Columbia University Medical Center, New York, NY, United States
[4]Washington University in St Louis, St Louis, MO, United States

Corresponding Author:
Walter Wai-Yip Chan, MD, MPH
Brigham and Women's Hospital
75 Francis Street
Boston, MA, 02115
United States
Phone: 1 617 732 6389
Email: wwchan@bwh.harvard.edu

Related Article:

Correction of: https://www.jmir.org/2019/9/e13345/

(J Med Internet Res 2021;23(3):e29042) doi:10.2196/29042

In "Association Between Institutional Social Media Involvement and Gastroenterology Divisional Rankings: Cohort Study" (J Med Internet Res 2019;21(9):e13345) the authors noted one error.

In the originally published paper, the name of author Loren Galler Rabinowitz was displayed incorrectly as surname "Galler Rabinowitz" and given name "Loren." This has been changed to the correct format of surname "Rabinowitz" and given names "Loren Galler." Accordingly, the citation for this author will be corrected from "Galler Rabinowitz L" to "Rabinowitz LG."

The correction will appear in the online version of the paper on the JMIR Publications website on March 29, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 23.03.21; this is a non–peer-reviewed article;accepted 23.03.21; published 29.03.21.

Please cite as:
Chiang AL, Rabinowitz LG, Kumar A, Chan WWY
Correction: Association Between Institutional Social Media Involvement and Gastroenterology Divisional Rankings: Cohort Study
J Med Internet Res 2021;23(3):e29042
URL: https://www.jmir.org/2021/3/e29042
doi:10.2196/29042
PMID:33780348

Corrigenda and Addenda

Correction: Associations Between Digital Health Intervention Engagement, Physical Activity, and Sedentary Behavior: Systematic Review and Meta-analysis

Matthew Mclaughlin[1,2,3,4], BSc; Tessa Delaney[1,2,3,4], BND; Alix Hall[1,2,3,4], PhD; Judith Byaruhanga[1,2,3,4], MPH; Paul Mackie[5,6], MSc; Alice Grady[1,2,3,4], PhD; Kathryn Reilly[1,2,3,4], PhD; Elizabeth Campbell[1,2,3,4], PhD; Rachel Sutherland[1,2,3,4], PhD; John Wiggers[1,2,3,4], PhD; Luke Wolfenden[1,2,3,4], PhD

[1]School of Medicine and Public Health, University of Newcastle, Callaghan, Australia

[2]Hunter New England Population Health, Wallsend, Australia

[3]Hunter Medical Research Institute, New Lambton Heights, Australia

[4]Priority Research Centre for Heath Behaviour, University of Newcastle, Callaghan, Australia

[5]School of Health Sciences and Priority Research Centre for Stroke and Brain Injury, University of Newcastle, Callaghan, Australia

[6]Centre for Research Excellence in Stroke Recovery and Rehabilitation, Florey Institute of Neuroscience, Melbourne, Australia

Corresponding Author:
Matthew Mclaughlin, BSc
School of Medicine and Public Health
University of Newcastle
University Drive
Callaghan, 2308
Australia
Phone: 61 02 4924 6477
Email: Matthew.Mclaughlin1@health.nsw.gov.au

Related Article:

Correction of: https://www.jmir.org/2021/2/e23180/

(J Med Internet Res 2021;23(3):e29094) doi:10.2196/29094

In "Associations Between Digital Health Intervention Engagement, Physical Activity, and Sedentary Behavior: Systematic Review and Meta-analysis" (J Med Internet Res 2021;23(2):e23180) a character display error was noted in 3 tables.

The "gamma" symbol (γ) was not properly rendered in 5 places in the paper due to an XML conversion error.

In Table 4, row "Rebar et al," column "Association":

=0.51 (95% CI −1.77 to 2.72); P>.05

has been corrected to:

γ=0.51 (95% CI −1.77 to 2.72); P>.05

In Table 5, row "Rebar et al, Time," column "Association":

=2.33 (95% CI 0.09 to 4.64); P<.05

has been corrected to:

γ=2.33 (95% CI 0.09 to 4.64); P<.05

and:

=0.51 (95% CI −1.77 to 2.72); P>.05

has been corrected to:

γ=0.51 (95% CI −1.77 to 2.72); P>.05

In Table 5, row "Rebar et al, Logins," column "Association":

=3.18 (95% CI 1.15 to 5.07); P<.05

has been corrected to:

γ=3.18 (95% CI 1.15 to 5.07); P<.05

and:

=2.04 (95% CI 0.29 to 3.84); P<.05

has been corrected to:

γ=2.04 (95% CI 0.29 to 3.84); P<.05

The correction will appear in the online version of the paper on the JMIR Publications website on March 29, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 25.03.21; this is a non–peer-reviewed article;accepted 25.03.21; published 29.03.21.

Please cite as:
Mclaughlin M, Delaney T, Hall A, Byaruhanga J, Mackie P, Grady A, Reilly K, Campbell E, Sutherland R, Wiggers J, Wolfenden L Correction: Associations Between Digital Health Intervention Engagement, Physical Activity, and Sedentary Behavior: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(3):e29094
URL: https://www.jmir.org/2021/3/e29094
doi:10.2196/29094
PMID:33780346

Corrigenda and Addenda

Correction: Opportunities and Challenges for Digital Social Prescribing in Mental Health: Questionnaire Study

Shivani Patel[1], MRCPsych; Gerry Craigen[2,3], MD; Mariana Pinto da Costa[1,4,5], MD; Becky Inkster[6,7], DPhil

[1]South London and Maudsley NHS Trust, London, United Kingdom

[2]Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[3]Centre for Mental Health, University Health Network – Toronto General, Toronto, ON, Canada

[4]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[5]Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal

[6]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

[7]Finance and Economics Programme, The Alan Turing Institute, London, United Kingdom

Corresponding Author:
Shivani Patel, MRCPsych
South London and Maudsley NHS Trust
Denmark Hill
London,
United Kingdom
Phone: 44 3228 6000 ext 1234
Email: shivaninpatel183@gmail.com

Related Article:

Correction of: https://www.jmir.org/2021/3/e17438/

(J Med Internet Res 2021;23(3):e29041) doi:10.2196/29041

In "Opportunities and Challenges for Digital Social Prescribing in Mental Health: Questionnaire Study" (J Med Internet Res 2021;23(3):e17438) the authors noted three errors.

In the originally published article, affiliation 1 was not applied to author Mariana Pinto da Costa, and the affiliation *"Centre for Mental Health, University Health Network – Toronto General, Toronto, ON, Canada"* was not added for author Gerry Craigen.

The author's affiliation section was originally published as follows:

> *Shivani Patel[1], MRCPsych; Gerry Craigen[2], MD; Mariana Pinto da Costa[3,4], MD; Becky Inkster[5,6], DPhil*
>
> *[1]South London and Maudsley NHS Trust, London, United Kingdom*
>
> *[2]Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, ON, Canada*
>
> *[3]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom*
>
> *[4]Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal*
>
> *[5]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom*
>
> *[6]Finance and Economics Programme, The Alan Turing Institute, London, United Kingdom*

The corrected author's affiliation section appears as follows:

> *Shivani Patel[1], MRCPsych; Gerry Craigen[2,3], MD; Mariana Pinto da Costa[1,4,5], MD; Becky Inkster[6,7], DPhil*
>
> *[1]South London and Maudsley NHS Trust, London, United Kingdom*
>
> *[2]Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, ON, Canada*
>
> *[3]Centre for Mental Health, University Health Network – Toronto General, Toronto, ON, Canada*
>
> *[4]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom*
>
> *[5]Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal*
>
> *[6]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom*

[7]*Finance and Economics Programme, The Alan Turing Institute, London, United Kingdom*

In the "Data Analysis" section, a correction has been made to clarify the author order referenced in the text.

The following sentence appeared in the originally published manuscript:

> *The first author (SP) coded all the material, and the second author (MP) reviewed all the data to ensure the consistency and credibility of the coding and grouping [18].*

This sentence has been corrected to:

> *The first author (SP) coded all the material, and the third author (MP) reviewed all the data to ensure the consistency and credibility of the coding and grouping [18].*

The correction will appear in the online version of the paper on the JMIR Publications website on March 29, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 23.03.21; this is a non–peer-reviewed article;accepted 23.03.21; published 29.03.21.

Please cite as:
Patel S, Craigen G, Pinto da Costa M, Inkster B
Correction: Opportunities and Challenges for Digital Social Prescribing in Mental Health: Questionnaire Study
J Med Internet Res 2021;23(3):e29041
URL: https://www.jmir.org/2021/3/e29041
doi:10.2196/29041
PMID:33780347

Corrigenda and Addenda

Correction: Measurement of Digital Literacy Among Older Adults: Systematic Review

Sarah Soyeon Oh[1], PhD; Kyoung-A Kim[2], PhD; Minsu Kim[3]; Jaeuk Oh[3]; Sang Hui Chu[1], PhD; JiYeon Choi[1], PhD

[1]Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, Seoul, Republic of Korea

[2]Department of Nursing, Yeoju Institute of Technology, Yeoju, Gyeonggi-do, Republic of Korea

[3]College of Nursing, Yonsei University, Seoul, Republic of Korea

Corresponding Author:
JiYeon Choi, PhD
Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University
50 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 3301
Fax: 82 2 2227 8303
Email: jychoi610@yuhs.ac

Related Article:

Correction of: https://www.jmir.org/2021/2/e26145/

(J Med Internet Res 2021;23(3):e28211) doi:10.2196/28211

In the originally published paper, the footnotes under Table 3 were incorrect.

The footnotes originally appeared as follows:

[b]*O: not included in the questionnaire*

[c]*X: included in the questionnaire*

These have now been corrected to the following:

[b]*O: included in the questionnaire*

[c]*X: not included in the questionnaire*

The correction will appear in the online version of the paper on the JMIR Publications website on March 03, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 24.02.21; this is a non–peer-reviewed article;accepted 25.02.21; published 03.03.21.

Please cite as:
Oh SS, Kim KA, Kim M, Oh J, Chu SH, Choi J
Correction: Measurement of Digital Literacy Among Older Adults: Systematic Review
J Med Internet Res 2021;23(3):e28211
URL: https://www.jmir.org/2021/3/e28211
doi:10.2196/28211
PMID:33657006

Corrigenda and Addenda

Correction: Effects of COVID-19 Emergency Alert Text Messages on Practicing Preventive Behaviors: Cross-sectional Web-Based Survey in South Korea

Minjung Lee[1,2], PhD; Myoungsoon You[1], MPH, PhD

[1]Department of Public Health Sciences, Graduate School of Public Health, Seoul National University, Seoul, Republic of Korea

[2]Office of Dental Education, School of Dentistry, Seoul National University, Seoul, Republic of Korea

Corresponding Author:
Myoungsoon You, MPH, PhD
Department of Public Health Sciences, Graduate School of Public Health
Seoul National University
Gwanak-ro 1, Gwanak-gu
Seoul, 08826
Republic of Korea
Phone: 82 880 2773
Email: msyou@snu.ac.kr

Related Article:

Correction of: https://www.jmir.org/2021/2/e24165/

(J Med Internet Res 2021;23(3):e28660) doi:10.2196/28660

In "Effects of COVID-19 Emergency Alert Text Messages on Practicing Preventive Behaviors: Cross-sectional Web-Based Survey in South Korea" (J Med Internet Res 2021;23(2):e24165) the authors noted an error in authorship.

In the originally published article, the order of authors was incorrectly listed as follows:

Myoungsoon You, Minjung Lee

The authorship list has been corrected to:

Minjung Lee, Myoungsoon You

The correction will appear in the online version of the paper on the JMIR Publications website on March 18, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 09.03.21; this is a non–peer-reviewed article;accepted 11.03.21; published 18.03.21.

Please cite as:
Lee M, You M
Correction: Effects of COVID-19 Emergency Alert Text Messages on Practicing Preventive Behaviors: Cross-sectional Web-Based Survey in South Korea
J Med Internet Res 2021;23(3):e28660
URL: https://www.jmir.org/2021/3/e28660
doi:10.2196/28660
PMID:33735093

Corrigenda and Addenda

Correction: COVID-19 Discourse on Twitter in Four Asian Countries: Case Study of Risk Communication

Sungkyu Park[1*], PhD; Sungwon Han[2*], MSc; Jeongwook Kim[2*], MSc; Mir Majid Molaie[2], BSc; Hoang Dieu Vu[3], MSc; Karandeep Singh[1], PhD; Jiyoung Han[2], PhD; Wonjae Lee[2], PhD; Meeyoung Cha[1,2], PhD

[1]Data Science Group, Institute for Basic Science, Daejeon, Republic of Korea

[2]Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

[3]Electrical and Electronic Engineering, Phenikaa University, Hanoi, Vietnam

[*]these authors contributed equally

Corresponding Author:
Meeyoung Cha, PhD
Data Science Group
Institute for Basic Science
55, Expo-ro, Yuseong-gu
Daejeon, 34126
Republic of Korea
Phone: 82 428788114
Fax: 82 428788079
Email: meeyoung.cha@gmail.com

Related Article:

Correction of: https://www.jmir.org/2021/3/e23272/

(J Med Internet Res 2021;23(3):e28926) doi:10.2196/28926

In "COVID-19 Discourse on Twitter in Four Asian Countries: Case Study of Risk Communication" (J Med Internet Res 2021;23(3):e23272) the authors noted one error.

In the originally published article, the caption of Figure 10 was incorrectly described. It read as follows:

> *Daily topic trends in Vietnam based on the number of tweets.*

This has been corrected to:

> *Daily topic trends in India based on the number of tweets.*

The correction will appear in the online version of the paper on the JMIR Publications website on March 29, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 18.03.21; this is a non–peer-reviewed article;accepted 23.03.21; published 29.03.21.

Please cite as:
Park S, Han S, Kim J, Molaie MM, Vu HD, Singh K, Han J, Lee W, Cha M
Correction: COVID-19 Discourse on Twitter in Four Asian Countries: Case Study of Risk Communication
J Med Internet Res 2021;23(3):e28926
URL: https://www.jmir.org/2021/3/e28926
doi:10.2196/28926
PMID:33780345

Original Paper

Exploring Usage of COVID Coach, a Public Mental Health App Designed for the COVID-19 Pandemic: Evaluation of Analytics Data

Beth K Jaworski[1*], PhD; Katherine Taylor[1*], MPH, PsyD; Kelly M Ramsey[1*]; Adrienne Heinz[1,2*], PhD; Sarah Steinmetz[1*], PhD; Ian Pagano[3*], PhD; Giovanni Moraja[4*]; Jason E Owen[1*], MPH, PhD

[1]National Center for PTSD, Dissemination & Training Division, US Department of Veterans Affairs, Menlo Park, CA, United States

[2]School of Medicine, Stanford University, Stanford, CA, United States

[3]University of Hawaii Cancer Center, Honolulu, HI, United States

[4]Vertical Design, LLC, Berkeley, CA, United States

[*]all authors contributed equally

Corresponding Author:
Beth K Jaworski, PhD
National Center for PTSD, Dissemination & Training Division
US Department of Veterans Affairs
795 Willow Road
Menlo Park, CA, 94025
United States
Phone: 1 650 308 9437
Email: beth.jaworski@va.gov

Abstract

Background: The COVID-19 pandemic has significantly impacted mental health and well-being. Mobile mental health apps can be scalable and useful tools in large-scale disaster responses and are particularly promising for reaching vulnerable populations. COVID Coach is a free, evidence-informed mobile app designed specifically to provide tools and resources for addressing COVID-19–related stress.

Objective: The purpose of this study was to characterize the overall usage of COVID Coach, explore retention and return usage, and assess whether the app was reaching individuals who may benefit from mental health resources.

Methods: Anonymous usage data collected from COVID Coach between May 1, 2020, through October 31, 2020, were extracted and analyzed for this study. The sample included 49,287 unique user codes and 3,368,931 in-app events.

Results: Usage of interactive tools for coping and stress management comprised the majority of key app events (n=325,691, 70.4%), and the majority of app users tried a tool for managing stress (n=28,009, 58.8%). COVID Coach was utilized for ≤3 days by 80.9% (n=34,611) of the sample whose first day of app use occurred within the 6-month observation window. Usage of the key content in COVID Coach predicted returning to the app for a second day. Among those who tried at least one coping tool on their first day of app use, 57.2% (n=11,444) returned for a second visit; whereas only 46.3% (n=10,546) of those who did not try a tool returned (P<.001). Symptoms of anxiety, depression, and posttraumatic stress disorder (PTSD) were prevalent among app users. For example, among app users who completed an anxiety assessment on their first day of app use (n=4870, 11.4% of users), 55.1% (n=2680) reported levels of anxiety that were moderate to severe, and 29.9% (n=1455) of scores fell into the severe symptom range. On average, those with moderate levels of depression on their first day of app use returned to the app for a greater number of days (mean 3.72 days) than those with minimal symptoms (mean 3.08 days; t_1=3.01, P=.003). Individuals with significant PTSD symptoms on their first day of app use utilized the app for a significantly greater number of days (mean 3.79 days) than those with fewer symptoms (mean 3.13 days; t_1=2.29, P=.02).

Conclusions: As the mental health impacts of the pandemic continue to be widespread and increasing, digital health resources, such as apps like COVID Coach, are a scalable way to provide evidence-informed tools and resources. Future research is needed to better understand for whom and under what conditions the app is most helpful and how to increase and sustain engagement.

(J Med Internet Res 2021;23(3):e26559) doi:10.2196/26559

KEYWORDS

COVID-19; coronavirus; mobile app; mHealth; digital health; mental health; public mental health; stress; coping; public health; app

Introduction

Impact of COVID-19 on Mental Health and Well-Being

In the United States, the COVID-19 pandemic has led to over 500,000 deaths, millions of job losses, and disruption of nearly every aspect of daily life. COVID-19 has also negatively impacted mental health and well-being globally [1-3]. One-third of American adults report a high level of psychological distress due to the pandemic [4].

Several studies now indicate that an unprecedented mental health crisis is underway. In a poll conducted by Harris [5] on behalf of the American Psychological Association, nearly 8 in 10 adults said the pandemic is a significant source of stress in their lives. The prevalence of depression symptoms among adults in the United States has risen from 8.5% of the population prior to the COVID-19 pandemic to 27.8% in the midst of the pandemic [6]. Researchers from the US Centers for Disease Control and Prevention found that 40% of respondents of a survey administered in June 2020 endorsed at least one adverse mental or behavioral health condition including symptoms of depression, anxiety, posttraumatic stress, or having started or increased substance use to cope with stress or emotions related to COVID-19. Over 10% of respondents reported seriously considering suicide in the previous 30 days [7]. Furthermore, there appears to be a bidirectional relationship between COVID-19 and psychiatric disorders, such that having a psychiatric disorder is associated with a greater likelihood of contracting COVID-19, and contracting COVID-19 is associated with an increased risk of receiving a psychiatric diagnosis [8].

Digital Mental Health as a Strategy for Addressing the Mental Health Impact of COVID-19

Digital mental health options are needed to help address the mental health effects of COVID-19 as well as the secondary impacts of the pandemic, such as fear of contracting the virus, financial stress related to job loss, loss of childcare, or the need to balance work with remote education. Mobile mental health apps are a promising strategy for addressing mental health impacts of the pandemic because of their potential scalability, reach, and utility, particularly during a time when in-person care may not be accessible due to social distancing and safety regulations. High-quality, accessible, and sustainable apps have been identified as part of an integrated "blueprint" for digital mental health services during the pandemic [9]. They may be a particularly useful tool for reaching a large number of individuals from highly impacted populations at risk for posttraumatic stress disorder (PTSD) or other mental health conditions, including those who have contracted COVID-19 and frontline health care workers [10].

Apps are a particularly appealing medium because of their potential reach. Individuals rarely turn off mobile devices [11], making apps available 24/7. Additionally, in the United States, 81% of adults own smartphones, with few differences among sociodemographic groups [12]. This reach is important because the pandemic has a disproportionate and complex impact on Black, Indigenous, people of color (BIPOC), people from low-income backgrounds, and women [13], and it is clear that vulnerable groups are at greater risk for behavioral and mental health consequences [6,7,14]. Systemic disadvantage with respect to social determinants of health, such as lack of internet access and reduced educational opportunities, has been associated with increased COVID-19 mortality rates [15]. Free, evidence-informed apps, such as COVID Coach, that are developed by government or not-for-profit entities and made specifically to address such systemic barriers, can contribute to a digital mental health safety net for vulnerable individuals. Beyond the ability to reach many people, apps have been shown to be useful adjunctive resources for a range of mental health concerns, including anxiety and depression [16] and PTSD [17].

Creation of the COVID Coach App

In response to the anticipated mental health impact of the COVID-19 pandemic, and as part of the Veterans Affairs' (VA) "Fourth Mission" to help during times of national emergencies and support public health, the National Center for PTSD created COVID Coach (Multimedia Appendix 1). COVID Coach is a free, publicly available mental health app designed to help people cope with stress, find resources, and track mental health over time. It is intended to be simple to use, does not require an internet connection or data plan to access primary content, and all recommended activities and resources are low in cost or free to users. COVID Coach is one of only a few public mental health apps available for specifically addressing mental health concerns stemming from or exacerbated by COVID-19, and it is the newest in a suite of free mental health apps designed to support mental health [18,19].

COVID Coach is based upon the model of the empirically supported PTSD Coach app [20], which has been identified as a potential approach for the behavioral and mental health impact of COVID-19 [21]. COVID Coach provides app users with many of the features of PTSD Coach, including tools for coping with challenging situations and managing stress, psychoeducation, tracking of mental health symptoms, and quick access to support networks and crisis resources. COVID Coach also provides symptom management tools adapted for life during the pandemic (eg, sleep struggles; isolation; stress; sadness; and indoor, socially distanced activities), goal-setting, and over 50 unique psychoeducational topics about managing COVID-19–related concerns (ie, staying well, staying balanced, staying together, staying safe, and staying healthy). The app was released at the end of April 2020 and has been promoted as part of the VA's response to the pandemic and highlighted as an important resource [22].

Evaluating COVID Coach in the Context of a Public Health Disaster

Mobile mental health apps can be useful tools in large-scale disaster responses [23], and their use has been indicated

specifically within the context of the COVID-19 pandemic (eg, [24,25]). However, the utility of standalone apps "in the wild" can be limited by poor engagement and high attrition (eg, [26,27].) A host of challenges renders it difficult to conduct formal research and evaluation on disaster mental health interventions and resources [28]. Accordingly, there is often insufficient data on when, how, and why individuals utilize disaster mental health resources to help guide policy and budgetary allocation. Although COVID Coach has been well received in the general population, usage of the app, particularly the key content areas, and retention have not yet been formally evaluated.

Objective

This study utilized anonymous mobile analytics data to characterize the overall usage of an app designed specifically to provide tools and resources for addressing COVID-19–related stress, explore retention and return usage, and assess whether the app was reaching individuals that may benefit from mental health resources. Three key aims guided the study: (1) describe general usage trends between May 1, 2020, and October 31, 2020 (a key period of time during the pandemic), and identify how frequently specific types of key app content were used (ie, coping tools, psychoeducation, self-assessments, and accessing resources); (2) explore usage patterns, with a particular focus on understanding how usage of key content on the first day of use may be related to return use and retention; and (3) characterize baseline mental health and well-being among COVID Coach users.

Methods

COVID Coach Mobile App Description

COVID Coach, available for Android [29] and iOS [30], is an app designed specifically for the COVID-19 pandemic to provide users with interactive, evidence-informed tools for coping with stress and anxiety, information about how to stay well, stay connected, and navigate challenges, self-monitoring mental health symptoms and goals, and resources to discover and connect with various types of verified and vetted support. The app can be used independently or in conjunction with professional mental health care but is not a replacement for therapy. Users are not required to create an account or log in to access any of the content, and the app is fully compatible with assistive software technologies (eg, VoiceOver or TalkBack).

Mobile Analytics Data

COVID Coach collects anonymous information about app use for the purposes of quality improvement. Fully nonidentifying, anonymous, and encrypted event sequences were stored using JavaScript Object Notation (JSON) format on a remote GovCloud server that meets VA security and privacy requirements. Data are accessible from VA App Connect software, which has been approved for use under the VA's Technical Reference Model [31]. Upon first launch of the app, a unique, randomly generated 32-character (256-bit) code is assigned to that particular app installation. Completely anonymous usage data, such as screens selected, button presses, and other nonidentifying patterns, are collected and associated with this install code. Install codes serve as a proxy for app users since the unique identity of each app user cannot be determined. Each in-app event contains a timestamp (in Coordinated Universal Time [UTC]) that corresponds to when the event actually occurred, but data are only transmitted to the server when the app is in use and connected to Wi-Fi or utilizing a data plan.

Procedures

For the purpose of this study, mobile analytics data with timestamps between May 1, 2020, and October 31, 2020, were extracted from the research server on November 4, 2020. Between May 1 and October 31, 3,368,931 in-app related events were captured (Android: n=847,260; iOS: n=2,521,612) across 49,297 unique install codes (Android: n=12,938; iOS: n=36,359).

Measures

App Use Metrics

Daily active users and monthly active users were measured by the total number of app users that used COVID Coach on a given day or at least once within a given month. Overall, frequencies for key content usage were computed for each of the four key sections in the app: *Manage Stress* (tried a tool), *Learn* (viewed a learn topic), *Mood Check* (created and rated a goal or completed an assessment), and *Find Resources* (viewed at least one specific subsection within *Find Resources*). These frequencies were computed for all key events and for all app users that had activity during the observation window (May 1, 2020, through October 31, 2020). Based on a rationale similar to Kwasny and colleagues [32], we decided a priori that frequency of use within the observation window would be measured in terms of unique days of use, rather than sessions or visits because of the variability in establishing the end of an app session, within and across platforms. Additionally, all app users were categorized according to whether their first day of app use occurred during the observation window (first-time users) or prior to the start of the observation window. Thus, all analyses related to distinct days of app use, return usage, and patterns of usage by day of use focused only on app events associated with first-time users. Among all first-time users, distinct days of app use within the observation window were calculated, as well as retention days (the number of days between the first day of use and the last day of use) and the number of days between the first day of use and the second day of use (for all individuals who used the app for at least 2 distinct days). For each first-time user, completion of tasks within each of the four key content areas were totaled, by each distinct day of use. First-time users who completed one or more assessments on their first day of app use were identified as "baseline" assessment completers.

In-App Assessments

Four assessments are available within the *Mood Check* section of COVID Coach. These assessments can be accessed and taken at any time by app users.

The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) [33] is a measure to assess the feelings and functional aspects

of positive mental health. COVID Coach contains the 14-item version of the scale, with each item measured on a Likert-type scale ranging from 1 (“none of the time”) to 5 (“all of the time”). For each item, respondents are asked to consider how they have been feeling over the past 2 weeks. Total score is obtained by summing all items. Scores of less than 42 are indicative of low well-being [34]. The scale was found to be a valid and reliable tool for measuring mental well-being in diverse populations and across project types, and has adequate internal reliability (α=.89) [35].

The Generalized Anxiety Disorder-7 (GAD-7) [36] is a measure to screen for GAD and assess severity of GAD symptoms. The scale consists of 7 items, each measured on a Likert-type scale ranging from 0 (“not at all”) to 3 (“nearly every day”), and total score is obtained by summing all items. For each item, respondents are asked to consider how they have been feeling over the past 2 weeks. Anxiety symptom severity is categorized as: minimal (total score=0-4), mild (total score=5-9), moderate (total score=10-14), and severe (total score=15 or higher). The scale has acceptable internal reliability and good psychometric properties, including among general population samples [37].

The Patient Health Questionnaire-9 (PHQ-9) [38] is a measure to assess the severity of depression symptoms. The scale consists of 9 items, each measured on a Likert-type scale ranging from 0 (“not at all”) to 3 (“nearly every day”), and total score is obtained by summing all items. For each item, respondents are asked to consider how they have been feeling over the past 2 weeks. Depression symptom severity is categorized as: minimal (total score=0-4), mild (total score=5-9), moderate (total score=10-14), moderately severe (total score=15-19), and severe (total score=20 or higher). The scale has acceptable internal reliability (α=.86-.89) and overall sound psychometric properties across settings [39].

The Posttraumatic Stress Disorder Checklist (PCL-5) [40] is a measure to assess symptoms of PTSD. The scale consists of 20 items, each measured on a Likert-type scale ranging from 0 (“not at all”) to 4 (“extremely”), and total score is obtained by summing all items. In COVID Coach, the PCL-5 is administered with only a brief introduction, followed by the assessment items. For each item, respondents are asked to consider how they have been feeling over the past month. Initial research suggests that total scores of 31 to 33 (or higher) are indicative of probable PTSD. For this study, we use 33 as the cut-off for significant PTSD symptoms. The PCL-5 was found to be reliable and valid in both veteran [41] and civilian populations [42].

Analyses

SQLPro Studio (Hankinsoft Development, Inc) was used for all data preprocessing and extraction. SAS University Edition (SAS Institute) software in conjunction with Oracle’s VirtualBox were used for all data analyses. We calculated descriptive statistics for key content usage, retention, and baseline levels of mental health symptom severity and levels of well-being. Chi-square analyses were conducted to understand differences in returning to the app for a second day of use based on key content usage on the first day of app use and baseline mental health symptoms. We ran separate chi-square analyses for each predictor. Independent samples *t* tests were conducted to examine differences in total unique days of app use and total manage stress tools utilized among app users who completed an assessment on their first day of app use compared to those who did not. An analysis of variance (ANOVA) was conducted with a Tukey test for post hoc analysis to examine differences in baseline WEMWBS scores, by month, among users who completed a well-being assessment on their first day of app use. Regression analyses were conducted to examine the relationship between baseline mental health symptoms and unique days of app usage.

Results

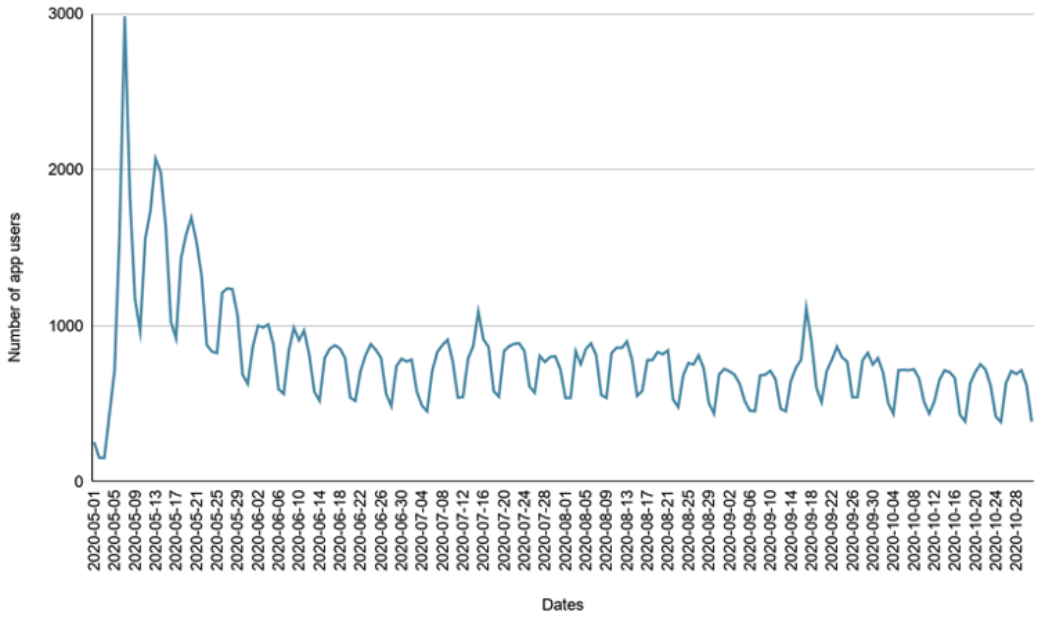
Reach and Reception

The app was released at the end of April 2020, and as of October 31, 2020, it has been downloaded 143,097 times. It is highly rated on both the Apple App Store (4.8 out 5 stars) and the Google Play Store (4.7 out of 5 stars). Users had the opportunity to provide written reviews along with star ratings. The majority of written reviews were overwhelmingly positive, with comments such as “Beautifully calming…,” “a necessity for our new normal,” “one of the best free apps I’ve found,” and “this is amazing… it has all you may need… mood trackers, resources, meditation-not too frilly, just important.” Notably, due to Google’s restrictions on mobile apps related to COVID-19 (including hiding certain results for apps among searches containing “COVID”), COVID Coach has been installed at a ratio of over 3:1 for iOS compared to Android mobile devices.

Daily and Monthly Active Users

The number of daily active users spiked in May 2020 (mean 1205.77, SD 615.70), shortly after the app’s release. The number of daily active users has leveled off but remained stable with average daily active users of 778.67 (SD 161.16), 752.03 (SD 152.07), 712.71 (SD 141.85), 682.83 (SD 150.83), and 611.35 (SD 128.60), respectively, during the months of June, July, August, September, and October 2020 (Figure 1). Although timestamp information is only captured in UTC, there appears to be a consistent, weekly pattern of usage such that the app is used more during the week than on weekends. The number of monthly active users followed a similar pattern as daily active users. The number of monthly active users peaked in May but remained steady through October 2020, at approximately 11,000 unique app users per month.

Figure 1. Daily active COVID Coach users, from May 1, 2020, through October 31, 2020.



Key Content Usage

Within the observation window (May 1, 2020, through October 31, 2020), there were 49,297 unique app users and 462,651 app events associated with the four key content areas (*Manage Stress*, *Learn*, *Mood Check*, and *Find Resources*). Table 1 provides an overview.

Of the four key sections of the app, the *Manage Stress* section, which contains tools for coping with stress and anxiety, was the most utilized. Across the observation window, there were 325,691 total tool use events (70.4% of all key events), among 28,009 unique install codes (56.82% of all unique install codes).

Within the *Manage Stress* section, app users can directly select individual tools from a list of all tools, or they can have a tool recommended to them by selecting from one of seven possible challenges related to the pandemic: (1) coping with stress, (2) feeling lonely, (3) creating space for myself, (4) feeling sad or hopeless, (5) handling anger and irritability, (6) navigating relationships, and (7) sleep struggles. Across all app users, 48.5% (n=23,885) selected at least one challenge. Among this group of app users, challenges related to coping with stress were the most commonly selected (n=12,696, 53.2%), followed by sleep struggles (n=9308, 39.0%) and feeling lonely (n=9153, 38.3%).

Table 1. Overall key content usage among all COVID Coach users within the observation window (between May 1, 2020, and October 31, 2020).

Key content area (specific in-app action)	Unique app users, n (%)[a]	Key events, n (%)[b]	Totals per app user, mean (SD); range
Manage Stress (tried at least one tool)	28,009 (56.8)	325,691 (70.4)	11.63 (30.33); 1-2124
Learn (viewed at least one topic)	10,124 (20.5)	52,123 (11.3)	5.15 (8.09); 1-267
Mood Check (entered and rated at least one goal or completed at least one assessment)	13,510 (27.4)	47,821 (10.3)	3.54 (12.52); 1-1008
Find Resources (viewed at least one specific subsection)	9418 (19.1)	37,016 (8.0)	3.93 (7.82); 1-329

[a]Total number of unique app users during the observation window=49,297. Percentage of total app users. Percentages in this column will not sum to 100% because app users could have completed actions across the four types of key content areas.

[b]Total key app events during the observation window=462,651. Percentage of total key app events.

Overall, the psychoeducation content within the *Learn* section of the app was consumed less frequently and by fewer users than the *Manage Stress* tools. Within the observation sample, there were 52,123 unique learn topic views (11.3% of all key events), among 20.5% of all app users (10,124/49,297). Four out of the five most viewed topics appeared in the first subsection within *Learn* (Staying Well).

In total, core activities within the *Mood Check* section comprised 10.3% (47,821/462,651) of all key events. Across the observation window, 27.4% of all app users (13,510/49,297) submitted at least one goal success rating or completed at least one of the four available assessments in the *Mood Check* section. There were 10,253 submitted goal success ratings across 2285 app users (4.6% of the total sample), and 37,568 completed assessments across 13,223 unique app users (26.8% of all users).

Across the eleven subsections within *Find Resources*, 19.1% (9418/49,297) of all app users viewed the resource pages 37,016 times across the observation window, representing 8% of all key events. Notably, although not the most frequently viewed subsection, *Crisis Resources* (which includes direct links to

phone lines, text support, and online chat for services such as the National Suicide Prevention Lifeline, Crisis Text Line, and Substance Abuse and Mental Health Services Administration's Helpline) was visited 3297 times (8.9% of all *Find Resources* visits) among 2131 unique users. Table 2 presents detailed information about key events within each of the four key sections.

Table 2. Detailed key content usage among all COVID Coach users within the observation window (between May 1, 2020, and October 31, 2020).

Key content area	Unique app users, n (%)[a]	Key app events, n (%)[b]
Manage Stress: top 5 most frequently used tools		
Ambient Sounds (an audio-only tool with no narration)	7041 (14.3)	18,493 (4.0)
Deep Breathing (an audio-guided exercise)	7870 (16.0)	16,011 (3.5)
Change Your Perspective (a tool with tips for how to replace negative thoughts with more helpful ones)	6721 (13.6)	12,480 (2.7)
Muscle Relaxation (an audio-guided exercise focused on relaxing distinct core body parts)	6037 (12.2)	11,599 (2.5)
Grounding (a tool with tips on how to stay connected to the present moment and surroundings)	5718 (11.6)	9767 (2.1)
Learn: top 5 most frequently viewed topics		
Prioritizing Yourself, Right Now	2131 (4.3)	2811 (0.6)
Managing Irritability	1856 (3.8)	2281 (0.5)
Finding Humor	1506 (3.1)	1817 (0.4)
Finding Calm	1442 (2.9)	1813 (0.4)
Sleep	1184 (2.4)	1511 (0.3)
Find Resources: top 3 most frequently viewed sections		
Finding Local Resources (for locating state-specific COVID-19 guidelines and information)	2927 (5.9)	6451 (1.4)
Meeting Your Needs (for basic needs support)	3176 (6.4)	6223 (1.3)
Mobile Apps to Support Mental Health (information about other free apps to support mental health)	2461 (5.0)	3748 (0.8)
Mood Check: completion of assessments, by type		
Track Mood (PHQ-9[c])	7698 (15.6)	11,732 (2.5)
Track Anxiety (GAD-7[d])	8115 (16.5)	11,649 (2.5)
Track Well-Being (WEMWBS[e])	6151 (12.5)	8860 (1.9)
Track PTSD[f] Symptoms (PCL-5[g])	3568 (7.2)	5327 (1.2)

[a]Total number of unique app users during the observation window=49,297. Percentage of total app users.

[b]Total key app events during the observation window=462,651. Percentage of total key app events.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]GAD-7: Generalized Anxiety Disorder-7.

[e]WEMWBS: Warwick-Edinburgh Mental Well-Being Scale.

[f]PTSD: posttraumatic stress disorder.

[g]PCL-5: Posttraumatic Stress Disorder Checklist-5.

Return Usage and Retention

Among the 49,297 app user install codes present in the observation window, 86.8% (n=42,783) of COVID Coach users had their first day of app use occur within the observation window. Thus, for the analyses presented in this section, usage patterns will be restricted to only the app users and events associated with those whose first day of app use occurred during the observation window.

Nearly half of COVID Coach users used the app for a single day (n=20,793, 48.6%), and an additional 32.3% (n=13,818) used the app for 2 or 3 days in total. Less than 2% of the sample (n=709) used the app for 15 or more distinct days (Table 3). On average, across all app users with ≥2 distinct days of app use (n=21,990), the number of days retained was 42.44 (SD 44.40, median 25, range 1-179). On average, the number of days between the first day of app use and the second day of app use was 14.65 (SD 24.52, median 4, range 1-176). Although the majority of app users who returned to the app for at least a second day returned within 14 days, there was variability, including users whose second day of use occurred over 90 days after the first (see Table 4 for a detailed analysis among users

whose first month of use occurred in May, June, or July so that returns within a 90-day or longer window could be examined).

Table 3. Total number of distinct days of COVID Coach use, by month of first app use.

Month of first app use	Frequency of users per distinct day, n (%)					
	1 day only	2 days	3 days	4-6 days	7-14 days	≥15 days
May	6573 (42.67)	3406 (22.11)	1891 (12.28)	2137 (13.87)	1049 (6.81)	348 (2.26)
June	2533 (44.26)	1205 (21.10)	693 (12.14)	770 (13.49)	372 (6.51)	137 (2.40)
July	2939 (48.79)	1246 (20.68)	640 (10.62)	737 (12.23)	362 (6.01)	100 (1.66)
August	3165 (51.31)	1267 (20.54)	648 (10.51)	698 (11.32)	314 (5.09)	76 (1.23)
September	2783 (53.66)	1119 (21.58)	548 (10.57)	503 (9.70)	190 (3.66)	43 (0.83)
October	2800 (65.25)	849 (19.79)	306 (7.13)	273 (6.36)	58 (1.35)	5 (0.12)
All	20,793 (48.6)	9092 (21.25)	4726 (11.05)	5118 (11.96)	2345 (5.48)	709 (1.66)

Table 4. Analysis of days between first and second app use among COVID Coach users with at least 2 distinct days of app use, by month of first use.

First month of app use	Users that returned at least once, n	First return within 7 days, n (%)	First return within 8-14 days, n (%)	First return within 15-30 days, n (%)	First return within 31-60 days, n (%)	First return within 61-90 days, n (%)	First return after more than 90 days, n (%)
May	8831	4956 (56.12)	1071 (12.13)	1107 (12.54)	870 (9.85)	391 (4.43)	436 (4.94)
June	3177	1786 (56)	403 (12.68)	457 (14.38)	298 (9.38)	118 (3.71)	115 (3.62)
July	3085	1922 (62.30)	385 (12.48)	366 (11.86)	232 (7.52)	130 (4.21)	50 (1.62)

Differential Day 2 Return Rates Based on Day 1 Key Content Usage

On both the first and second days of app use (see Table 5 for an overview of usage), many app users tried at least one tool within the *Manage Stress* section (46.80% [n=20,222] on the first day, 41.85% [n=9202] among individuals who returned to the app for a second day). Usage of the key content in COVID Coach predicted returning to the app for a second day.

Of those who tried at least one *Manage Stress* tool on their first day of app use, 57.2% (n=11,444) returned for a second visit; whereas only 46.3% (n=10,546) of those who did not try a tool returned (P<.001). Among those who viewed at least one *Learn* topic on their first day of app use, 58.8% (n=3292) returned for a second day of use; whereas only 50.3% (n=18,698) who did not view a learn topic returned (P<.001). With respect to the *Mood Check* section, 57.2% (n=4892) of app users that completed at least one goal rating or one assessment activity returned for a second day of use, compared to 50.0% (n=17,098) of users who did not complete any *Mood Check* activities (P<.001). Lastly, among app users who viewed at least one specific *Find Resources* subsection, 57.4% (n=3014) returned for a second day of app use, compared to only 50.6% (n=18,976) of users who did not view any resources returned (P<.001).

Additionally, usage patterns among individuals who completed an assessment on the first day of app use were significantly different than those who did not complete an assessment on their first day. On average, individuals who completed at least one assessment on their first day of app use utilized COVID Coach for more unique days within the observation window (mean 3.29 days, SD 5.44) compared to individuals who did not complete an assessment on the first day (mean 2.66 days, SD 4.37; P<.001). Similarly, individuals who completed at least one assessment on their first day of app use utilized, on average, significantly more *Manage Stress* tools within the observation window (mean 9.2 tools, SD 24.6) compared to individuals who did not complete an assessment on the first day (mean 5.8 tools, SD 19.3; P<.001).

Table 5. Comparison of key content area usage, by first and second day of app use.

Number of key content areas accessed	App users	
	First day of app use (n=42,783), n (%)	Second day of app use (n=21,990), n (%)
All four key areas		
Completed at least one action within all four key content areas	650 (1.5)	192 (0.9)
Two to three key areas		
Manage Stress (with one or two other key areas; tried at least one tool and completed another action within one or two other key areas)	7953 (18.6)	3143 (14.3)
Two or three key areas (excluding Manage Stress; completed at least one action within two or more of the Learn, Mood Check, or Find Resources sections)	1129 (2.6)	405 (1.8)
One key area		
Manage Stress only (only tried at least one tool)	11,419 (26.7)	5867 (26.7)
Mood Check only (only completed at least one goal rating or assessment)	2677 (6.3)	1355 (6.2)
Find Resources only (only viewed at least one resource subsection)	1196 (2.8)	660 (3.0)
Learn only (only viewed at least one learn topic)	805 (1.9)	485 (2.2)
No key area actions		
Did not complete an action within any of the four key areas	16,954 (39.6)	9883 (44.9)

Characterizing Baseline Mental Health Among COVID Coach Users

Baseline well-being among individuals using COVID Coach appeared to be relatively low and decreased over time. Among app users who completed a WEMWBS assessment on their first day of app use (n=3558, 8.32% of all users whose first day of app use occurred during the observation window), average well-being scores, by month, were all less than 42, which has been used as a cut-off to identify low well-being [34]. These average baseline scores decreased over time, with app users who completed their first WEMWBS on their first day of app use in September 2020 (n=416; mean 38.7, SD 0.04) or October 2020 (n=341; mean 38.1, SD 9.50) demonstrating significantly lower average well-being scores than app users who completed their first WEMWBS on their first day of using the app in May 2020 (n=1361; mean 41.2, SD 9.65).

Symptoms of anxiety, depression, and PTSD were prevalent among app users. For all app users who completed a GAD-7 assessment on their first day of app use (n=4870; 11.4% of users), 12.8% (n=625) had scores suggesting minimal anxiety (total score=0-4), 32.1% (n=1565) endorsed mild levels of anxiety (total score=5-9), 25.2% (n=1225) indicated moderate levels of anxiety (total score=10-14), and 29.9% (n=1455) of scores fell into the severe symptom range (total score=15 or higher).

Among app users who completed a PHQ-9 on their first day of app use (n=4548, 10.6% of users), 16.5% (n=749) had scores suggesting minimal depression (total score=0-4), 28.9% (n=1312) endorsed mild levels of depression (total score=5-9), 25.0% (n=1136) indicated moderate levels of depression (total score=10-14), 17.5% (n=795) endorsed moderately severe levels of depression (total score=15-19), and 12.2% (n=556) of scores fell into the severe symptom range (total score=20 or higher).

Unlike the GAD-7 and the PHQ-9, the PCL-5 does not have symptom severity categorizations. However, among app users who completed a PCL-5 on their first day of app use (n=2064, 4.8% of users), the majority of individuals who completed the assessment (n=1234, 59.8%) had a total score ≥33, which is consistent with significant PTSD symptoms.

Baseline Mental Health Characteristics and Return Usage

Baseline PTSD symptoms predicted returning to the app for a second day. Among individuals with a baseline PCL-5 score of 33 or greater, 62.5% returned to the app for a second day of use, compared to only 56.4% of individuals with scores below 33 (P=.006). Neither symptom severity for anxiety or depression nor levels of well-being were predictive of return usage.

We conducted regression analyses to examine the relationship between baseline mental health symptoms and unique days of app usage. Depression and PTSD symptoms were predictive of the total number of unique days of app use. With respect to depression symptoms, we utilized the group with minimal symptoms as the reference group in comparison to those with mild, moderate, moderately severe, and severe symptoms. On average, those with moderate levels of depression on their first day of app use returned to the app for a greater number of days (mean 3.72 days) than those with minimal symptoms of depression (mean 3.08 days; t_1=3.01, P=.003). Individuals with mild, moderately severe, and severe depression did not significantly differ from the reference group. Although the

difference in usage between moderately severe and minimal symptom severity categories was not statistically significant, it was trending in the predicted direction. With respect to PTSD symptoms, individuals with baseline PCL-5 scores indicating significant PTSD symptoms utilized the app for a significantly greater number of days (mean 3.79 days) than those with subthreshold symptom levels (mean 3.13 days; t_1=2.29, P=.02).

Discussion

Principal Findings

This exploration of COVID Coach usage among the general population suggests that mobile apps may have the reach and accessibility necessary to be a useful medium for disseminating mental health information and resources to individuals experiencing stress related to the COVID-19 pandemic.

Between May 1, 2020, and October 31, 2020, the app was used by nearly 50,000 individuals, and daily active usage has remained steady over time. In addition to the total number of individuals reached, the key content within the app was utilized in over 450,000 instances. The stress management tools were most frequently used with over 28,000 users utilizing individual tools over 300,000 times. Further, each of the other three key content areas in the app were accessed tens of thousands of times by tens of thousands of users. This reach and scalability of COVID Coach across the general population is an example of how digital mental health tools can become successfully integrated into disaster response strategies. From a public mental health perspective (eg, [43]), being able to rapidly deploy evidence-informed tools and reliable health information via a free, accessible, and secure app is a way for the federal government to contribute to a digital mental health safety net and reduce barriers to accessing mental health resources.

Importantly, COVID Coach appears to be reaching individuals in need of mental health resources. On average, among app users who completed assessments during their first day of use, well-being was low, and the majority of individuals were indicating greater than minimal symptoms of anxiety, depression, and PTSD. Additionally, among app users who identified challenges they are facing, the majority reported difficulties with managing stress, troubles with sleep, and feelings of loneliness. We cannot determine if individuals utilizing COVID Coach are representative of the general population, but elevated levels of anxiety, depression, and posttraumatic stress are consistent with other research conducted during the pandemic [6,7,44]. Individuals with significant PTSD symptoms at baseline were more likely to return to the app for a second day of app use. On average, individuals with significant PTSD symptoms used the app for a greater number of days than those with subthreshold symptoms, and individuals with moderate depression used the app for more days than those with minimal symptoms. Greater usage among individuals with moderate depression symptoms is consistent with previous research [45].

Although overall app utilization data suggested considerable reach, engagement proved to be less consistent. Our analyses revealed that the majority of COVID Coach users (80.9%) utilized the app on ≤3 days. This finding is consistent with research indicating that self-management apps for mental health are often not used over extended periods of time [26,27,46]. However, as noted by Ng and colleagues [47], there is a need for more standardized reporting of measures related to user engagement and retention. The average number of retention days, as well as the number of days between days of use, suggest that the app may not be something that individuals use on a daily basis, but rather during moments of distress or need. This type of usage is consistent with the overall design of COVID Coach as a self-management tool, which does not provide any guidance on how often or when to use the app.

This research also provides some guidance on how engagement might be encouraged in future app versions. In general, app users that completed actions within the key content areas on the first day of app use were more likely to return for a second day of app use. More specifically, users that completed an assessment on the first day of app use were significantly more likely to use the app for a greater number of days and to use a greater number of stress management tools than app users who did not complete an assessment on the first day of app use. These findings suggest that finding ways to motivate users to complete actions within key areas on their first day of app use, particularly tools and assessments, may be one way to enhance engagement and retention. For example, having recommendations for a tool or assessment to try, easily accessible from the app home screen, may encourage users to try a specific in-app activity. Additionally, the onboarding sequence could include a few brief questions to help tailor in-app recommendations to the user's intentions and preferences, and guide them through the process of setting customized goals for using the features within the app most relevant to them. Lastly, finding ways to regularly disseminate and highlight new app content (eg, managing stress around prolonged distance learning, vaccine information) may encourage users to return to the app more frequently.

Limitations

Because COVID Coach does not collect any identifying information, we cannot say anything about the populations that we have reached, other than what we can characterize based upon in-app actions. Future research that permits collection of identifying information is needed, particularly given the disproportionate impact the pandemic has had on vulnerable groups of people. A Spanish version of COVID Coach has recently been released, and plans for data collection on app usage within Spanish-speaking populations are underway.

Additionally, we utilize the unique install codes as a proxy for an individual user. We assume that most individuals do not delete and reinstall the app multiple times. However, if an individual were to download COVID Coach on more than one mobile device, or delete it and reinstall, each of those installations would be assigned a unique install code, and would appear as a new user.

Although the app includes assessments for individuals to self-monitor well-being and symptoms of anxiety, depression, and PTSD, it is difficult to reliably measure change in these constructs via the app, due to the naturalistic nature of this study

and the changing landscape of the pandemic over time. It is important to highlight that even though a score of 33 or higher on the PCL-5 is suggestive of PTSD, the assessment questions in the app do not ask app users to respond to the questions while focusing on a particular traumatic incident, so caution in interpreting the meaning of these scores is warranted. Because the PCL-5 refers to “the stressful experience” in each item, in the context of COVID Coach, the PCL-5 may be capturing overall levels of distress. While desirable, we also did not have a way to measure other potential proxy variables of interest such as coping self-efficacy, perceived helpfulness of the app, improved opinions about mental health care, or reduction in stress related to enhanced support access, as these cannot be determined solely by in-app usage data.

Future research is needed to better understand who is interested in public mental health apps like COVID Coach, what their primary goals are for using the app, which outcomes are most useful in understanding engagement patterns, and how successful usage is defined. For example, someone may use the app only once, find the exact resource they need, and not use the app again, whereas someone else may be experiencing significant stress, use tools in moments of distress, and track mental health symptoms on a weekly basis. Findings from this type of research could be used to advance the science of mobile mental health and also be directly applied to a suite of publicly available apps that have been downloaded over 4 million times and are in widespread use across the VA, the largest health care organization in the United States.

Conclusions

As the mental health impacts of the pandemic continue to be widespread and increasing, digital health resources, such as apps like COVID Coach, are a scalable way to provide evidence-informed tools and resources. We believe that this is the first evaluation of a mobile mental health app designed specifically for use during the COVID-19 pandemic. This work shows that tens of thousands of people are accessing the app, with a particular focus on the tools for stress and coping. Such rapid uptake of a public mobile mental health app is unprecedented and signals perceived value. Specially, the findings from this evaluation suggest that apps may play a helpful role in providing mental health resources in the context of a public health disaster.

Future research should attempt to elucidate for whom and under what conditions the app is most helpful, and how to increase and sustain engagement. Additional areas of focus should include how to optimize the app for populations impacted by disparities related to mental health literacy, digital literacy, and stigma around mental health care. As noted by many mHealth (mobile health) scholars [48-50], there is no reason to believe that digital mental health care and blended options will disappear after the pandemic, so it is important to find strategies for increasing reach and optimizing for engagement within self-management tools. These strategies must also attend to issues of health inequities [48,49,51]. Due to the scale of the crisis, the pandemic may have opened the door to conversations about mental health, and apps may be a helpful first step in providing tools, accurate information, and connecting people with reliable resources. Those in government and nonprofit organizations may be able to provide these kinds of tools as a way to contribute to a digital mental health safety net and help alleviate mental health disparities.

Acknowledgments

This paper was not sponsored by any funder. The views expressed in this work are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

Authors' Contributions

All authors contributed to the conceptualization, writing, and editing of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COVID Coach screenshots.
[PDF File (Adobe PDF File), 2624 KB - jmir_v23i3e26559_app1.pdf]

References

1. Liu CH, Zhang E, Wong GTF, Hyun S, Hahm HC. Factors associated with depression, anxiety, and PTSD symptomatology during the COVID-19 pandemic: Clinical implications for U.S. young adult mental health. Psychiatry Res 2020 Aug;290:113172 [FREE Full text] [doi: 10.1016/j.psychres.2020.113172] [Medline: 32512357]
2. Rodríguez-Rey R, Garrido-Hernansaiz H, Collado S. Psychological Impact and Associated Factors During the Initial Stage of the Coronavirus (COVID-19) Pandemic Among the General Population in Spain. Front Psychol 2020;11:1540 [FREE Full text] [doi: 10.3389/fpsyg.2020.01540] [Medline: 32655463]
3. Torjesen I. Covid-19: Mental health services must be boosted to deal with "tsunami" of cases after lockdown. BMJ 2020 May 15;369:m1994. [doi: 10.1136/bmj.m1994] [Medline: 32417756]

4. Keeter S. A third of Americans experienced high levels of psychological distress during the coronavirus outbreak. Pew Research Center. 2020 May 7. URL: https://www.pewresearch.org/fact-tank/2020/05/07/a-third-of-americans-experienced-high-levels-of-psychological-distress-during-the-coronavirus-outbreak/ [accessed 2021-02-24]
5. Stress in America 2020: A national mental health crisis. American Psychological Association. 2020 Oct. URL: https://www.apa.org/news/press/releases/stress/2020/sia-mental-health-crisis.pdf [accessed 2021-02-24]
6. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of Depression Symptoms in US Adults Before and During the COVID-19 Pandemic. JAMA Netw Open 2020 Sep 01;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
7. Czeisler ME, Lane RI, Petrosky E, Wiley JF, Christensen A, Njai R, et al. Mental Health, Substance Use, and Suicidal Ideation During the COVID-19 Pandemic - United States, June 24-30, 2020. MMWR Morb Mortal Wkly Rep 2020 Aug 14;69(32):1049-1057 [FREE Full text] [doi: 10.15585/mmwr.mm6932a1] [Medline: 32790653]
8. Taquet M, Luciano S, Geddes JR, Harrison PJ. Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort studies of 62 354 COVID-19 cases in the USA. The Lancet Psychiatry 2021 Feb 09;8(2):130-140 [FREE Full text] [doi: 10.1016/S2215-0366(20)30462-4] [Medline: 33181098]
9. Balcombe L, De Leo D. An Integrated Blueprint for Digital Mental Health Services Amidst COVID-19. JMIR Ment Health 2020 Jul 22;7(7):e21718 [FREE Full text] [doi: 10.2196/21718] [Medline: 32668402]
10. Alexopoulos AR, Hudson JG, Otenigbagbe O. The Use of Digital Applications and COVID-19. Community Ment Health J 2020 Oct 30;56(7):1202-1203 [FREE Full text] [doi: 10.1007/s10597-020-00689-2] [Medline: 32734311]
11. Rainie L, Zickhur K. Americans' views on mobile etiquette. Pew Research Center. 2015 Jul 26. URL: https://www.pewresearch.org/internet/2015/08/26/americans-views-on-mobile-etiquette/#:~:text=Some%2092%25%20of%20U.S.%20adults,they%20rarely%20turn%20it%20off [accessed 2021-02-24]
12. Mobile fact sheet. Pew Research Center. 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-02-24]
13. Kantamneni N. The impact of the COVID-19 pandemic on marginalized populations in the United States: A research agenda. J Vocat Behav 2020 Jun;119:103439 [FREE Full text] [doi: 10.1016/j.jvb.2020.103439] [Medline: 32390658]
14. Double Jeopardy: COVID-19 and Behavioral Health Disparities for Black and Latino Communities in the U.S. Substance Abuse and Mental Health Services Administration (SAHMSA). URL: https://www.samhsa.gov/sites/default/files/covid19-behavioral-health-disparities-black-latino-communities.pdf [accessed 2021-02-24]
15. Dalsania AK, Fastiggi MJ, Kahlam A, Shah R, Patel K, Shiau S, et al. The Relationship Between Social Determinants of Health and Racial Disparities in COVID-19 Mortality. J Racial Ethn Health Disparities 2021 Jan 05 [FREE Full text] [doi: 10.1007/s40615-020-00952-y] [Medline: 33403652]
16. Mohr DC, Schueller SM, Tomasino KN, Kaiser SM, Alam N, Karr C, et al. Comparison of the Effects of Coaching and Receipt of App Recommendations on Depression, Anxiety, and Engagement in the IntelliCare Platform: Factorial Randomized Controlled Trial. J Med Internet Res 2019 Aug 28;21(8):e13609 [FREE Full text] [doi: 10.2196/13609] [Medline: 31464192]
17. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PTSD Coach app: A pilot randomized controlled trial with community trauma survivors. Psychol Trauma 2016 May 25;8(3):384-392. [doi: 10.1037/tra0000092] [Medline: 27046668]
18. Mobile Apps. VA National Center for PTSD. URL: https://www.ptsd.va.gov/appvid/mobile/index.asp [accessed 2021-02-24]
19. Mobile Mental Health. VA National Center for PTSD, Dissemination & Training Division. URL: https://www.myvaapps.com/ [accessed 2021-02-24]
20. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. J Consult Clin Psychol 2017 Mar;85(3):267-273. [doi: 10.1037/ccp0000163] [Medline: 28221061]
21. Meshick AB, Ryan L, Cullen T. Beyond PPE: Protecting Health Care Workers To Prevent A Behavioral Health Disaster. Health Affairs. 2020 Jun 4. URL: https://www.healthaffairs.org/do/10.1377/hblog20200603.842660/full/ [accessed 2021-02-24]
22. Huang L. Applications as Tools Amid the COVID-19 Pandemic. Journal of Hospital Librarianship 2020 Oct 09;20(4):376-390. [doi: 10.1080/15323269.2020.1823304]
23. Ruzek JI, Kuhn E, Jaworski BK, Owen JE, Ramsey KM. Mobile mental health interventions following war and disaster. Mhealth 2016;2:37 [FREE Full text] [doi: 10.21037/mhealth.2016.08.06] [Medline: 28293610]
24. Marshall JM, Dunstan DA, Bartik W. Treating Psychological Trauma in the Midst of COVID-19: The Role of Smartphone Apps. Front Public Health 2020;8:402 [FREE Full text] [doi: 10.3389/fpubh.2020.00402] [Medline: 33014955]
25. Wong AH, Pacella-LaBarbara ML, Ray JM, Ranney ML, Chang BP. Healing the Healer: Protecting Emergency Health Care Workers' Mental Health During COVID-19. Ann Emerg Med 2020 Oct;76(4):379-384 [FREE Full text] [doi: 10.1016/j.annemergmed.2020.04.041] [Medline: 32534830]
26. Baumel A, Muench F, Edan S, Kane JM. Objective User Engagement With Mental Health Apps: Systematic Search and Panel-Based Usage Analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]

27. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the Wild: Using Novel Data to Examine the Reach, Use, and Impact of PTSD Coach. JMIR Ment Health 2015;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
28. Grolnick WS, Schonfeld DJ, Schreiber M, Cohen J, Cole V, Jaycox L, et al. Improving adjustment and resilience in children following a disaster: Addressing research challenges. Am Psychol 2018 Apr;73(3):215-229. [doi: 10.1037/amp0000181] [Medline: 29446960]
29. US Department of Veterans Affairs. COVID Coach (Version 1.0). Google Play Store. 2020. URL: https://play.google.com/store/apps/details?id=gov.va.mobilehealth.ncptsd.covid&hl=en_US&gl=US [accessed 2021-02-24]
30. US Department of Veterans Affairs. COVID Coach (Version 1.0). App Store. 2020. URL: https://apps.apple.com/us/app/covid-coach/id1504705038 [accessed 2021-02-24]
31. VA Technical Reference Model. URL: https://www.oit.va.gov/Services/TRM/TRMHomePage.aspx [accessed 2020-11-20]
32. Kwasny MJ, Schueller SM, Lattie E, Gray EL, Mohr DC. Exploring the Use of Multiple Mental Health Apps Within a Platform: Secondary Analysis of the IntelliCare Field Trial. JMIR Ment Health 2019 Mar 21;6(3):e11572 [FREE Full text] [doi: 10.2196/11572] [Medline: 30896433]
33. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
34. Stewart-Brown S, Samaraweera PC, Taggart F, Kandala N, Stranges S. Socioeconomic gradients and mental health: implications for public health. Br J Psychiatry 2015 Jun 02;206(6):461-465. [doi: 10.1192/bjp.bp.114.147280] [Medline: 25792696]
35. Stewart-Brown S, Platt S, Tennant A. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a valid and reliable tool for measuring mental well-being in diverse populations and projects. J Epidemiol Community Health 2011;65:A38-A39 [FREE Full text] [doi: 10.1136/jech.2011.143586.86]
36. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
37. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
38. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
39. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med 2007 Nov;22(11):1596-1602 [FREE Full text] [doi: 10.1007/s11606-007-0333-y] [Medline: 17874169]
40. Weathers FW, Litz BT, Keane TM, Palmieri PA, Marx BP, Schnurr PP. The PTSD Checklist for DSM-5 (PCL-5). The National Center for PTSD. 2013. URL: https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp [accessed 2021-02-24]
41. Bovin MJ, Marx BP, Weathers FW, Gallagher MW, Rodriguez P, Schnurr PP, et al. Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (PCL-5) in veterans. Psychol Assess 2016 Nov;28(11):1379-1391. [doi: 10.1037/pas0000254] [Medline: 26653052]
42. Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Development and Initial Psychometric Evaluation. J Trauma Stress 2015 Dec;28(6):489-498. [doi: 10.1002/jts.22059] [Medline: 26606250]
43. Owen JE, Kuhn E, Jaworski BK, McGee-Vincent P, Juhasz K, Hoffman JE, et al. VA mobile apps for PTSD and related problems: public health resources for veterans and those who care for them. Mhealth 2018;4:28 [FREE Full text] [doi: 10.21037/mhealth.2018.05.07] [Medline: 30148141]
44. Rajkumar RP. COVID-19 and mental health: A review of the existing literature. Asian J Psychiatr 2020 Aug 10;52:102066 [FREE Full text] [doi: 10.1016/j.ajp.2020.102066] [Medline: 32302935]
45. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
46. Kozlov E, Bantum E, Pagano I, Walser R, Ramsey K, Taylor K, et al. The Reach, Use, and Impact of a Free mHealth Mindfulness App in the General Population: Mobile Data Analysis. JMIR Ment Health 2020 Nov 27;7(11):e23377 [FREE Full text] [doi: 10.2196/23377] [Medline: 33245289]
47. Ng MM, Firth J, Minen M, Torous J. User Engagement in Mental Health Apps: A Review of Measurement, Reporting, and Validity. Psychiatr Serv 2019 Jul 01;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
48. Figueroa CA, Aguilera A. The Need for a Mental Health Technology Revolution in the COVID-19 Pandemic. Front Psychiatry 2020 Jun 3;11:523 [FREE Full text] [doi: 10.3389/fpsyt.2020.00523] [Medline: 32581891]

49. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
50. Wind TR, Rijkeboer M, Andersson G, Riper H. The COVID-19 pandemic: The 'black swan' for mental health care and a turning point for e-health. Internet Interv 2020 Apr;20:100317 [FREE Full text] [doi: 10.1016/j.invent.2020.100317] [Medline: 32289019]
51. Smith JA, Judd J. COVID-19: Vulnerability and the power of privilege in a pandemic. Health Promot J Austr 2020 Apr;31(2):158-160 [FREE Full text] [doi: 10.1002/hpja.333] [Medline: 32197274]

Abbreviations

ANOVA: analysis of variance
BIPOC: Black, Indigenous, people of color
GAD-7: Generalized Anxiety Disorder-7
JSON: JavaScript Object Notation
mHealth: mobile health
PCL-5: Posttraumatic Stress Disorder Checklist-5
PHQ-9: Patient Health Questionnaire-9
PTSD: posttraumatic stress disorder
UTC: Coordinated Universal Time
VA: Veterans Affairs
WEMWBS: Warwick-Edinburgh Mental Well-Being Scale

Edited by C Basch; submitted 16.12.20; peer-reviewed by J Ruzek, J Ray; comments to author 04.01.21; revised version received 14.01.21; accepted 17.02.21; published 01.03.21.

Please cite as:
Jaworski BK, Taylor K, Ramsey KM, Heinz A, Steinmetz S, Pagano I, Moraja G, Owen JE
Exploring Usage of COVID Coach, a Public Mental Health App Designed for the COVID-19 Pandemic: Evaluation of Analytics Data
J Med Internet Res 2021;23(3):e26559
URL: https://www.jmir.org/2021/3/e26559
doi:10.2196/26559
PMID:33606656

Original Paper

Barriers to the Large-Scale Adoption of a COVID-19 Contact Tracing App in Germany: Survey Study

Annelies G Blom[1,2*], PhD; Alexander Wenz[1,2*], PhD; Carina Cornesse[1], PhD; Tobias Rettig[1], MA; Marina Fikel[1], MA; Sabine Friedel[1], MA; Katja Möhring[1,2], PhD; Elias Naumann[1], PhD; Maximiliane Reifenscheid[1], MA; Ulrich Krieger[1], PhD

[1]Collaborative Research Center 884 "Political Economy of Reforms", University of Mannheim, Mannheim, Germany

[2]School of Social Sciences, University of Mannheim, Mannheim, Germany

[*]these authors contributed equally

Corresponding Author:
Alexander Wenz, PhD
School of Social Sciences
University of Mannheim
A5, 6
Mannheim, 68131
Germany
Phone: 49 621 181 2298
Email: a.wenz@uni-mannheim.de

Abstract

Background: During the COVID-19 pandemic, one way to reduce further transmissions of SARS-CoV-2 is the widespread use of contact tracing apps. Such apps keep track of proximity contacts and warn contacts of persons who tested positive for an infection.

Objective: In this study, we analyzed potential barriers to the large-scale adoption of the official contact tracing app that was introduced in Germany on June 16, 2020.

Methods: Survey data were collected from 3276 adults during the week the app was introduced using an offline-recruited, probability-based online panel of the general adult population in Germany.

Results: We estimate that 81% of the population aged 18 to 77 years possess the devices and ability to install the official app and that 35% are also willing to install and use it. Potential spreaders show high access to devices required to install the app (92%) and high ability to install the app (91%) but low willingness (31%) to correctly adopt the app, whereas for vulnerable groups, the main barrier is access (62%).

Conclusions: The findings suggest a pessimistic view on the effectiveness of app-based contact tracing to contain the COVID-19 pandemic. We recommend targeting information campaigns at groups with a high potential to spread the virus but who are unwilling to install and correctly use the app, in particular men and those aged between 30 and 59 years. In addition, vulnerable groups, in particular older individuals and those in lower-income households, may be provided with equipment and support to overcome their barriers to app adoption.

(J Med Internet Res 2021;23(3):e23362) doi:10.2196/23362

KEYWORDS

digital health; mobile health; smartphone; mobile phone; app; digital technology; contact tracing; coronavirus; COVID-19; survey

Introduction

Since the outbreak of COVID-19, millions of people worldwide have been infected with SARS-CoV-2 [1]. In the absence of an effective vaccine or cure, societies across the globe are testing various combinations of measures to contain the spread of the virus [2]. Many countries have introduced lockdowns to reduce the number of new infections to a level that allows national health systems to treat all patients effectively despite the additional influx of seriously ill people [3].

While lockdowns have proven effective at reducing the spread of the virus, they have a major impact on the economy and

social life [4,5]. A less economically damaging measure is contact tracing, where persons who have been in close proximity of someone known to be infected are quarantined until they can be confirmed not infected (ie, tested negative) or, if confirmed infected (ie, tested positive), until they are not contagious anymore. In Germany, this task has been performed by officials in local public health departments, who through personal conversations with infected persons, have been collecting proximity contacts to inform them of their potential infection and to implement quarantines [6].

To grant some relief to this labor-intensive system and to account for regionally strewn sudden new outbreaks, scientists have been discussing app-based contact tracing as a supplementary measure [7]. Once installed on a smartphone, a contact tracing app warns users when they have been in close contact with an infected person and may advise them to go into quarantine and get tested for infection. Thus, if adopted widely, apps may allow for a more efficient tracing of infection chains.

On June 16, 2020, 141 days after the first diagnosis of COVID-19 in Germany [8], the federal government and the Robert Koch Institute (RKI) (ie, the German center for disease control and prevention) launched their official COVID-19 contact tracing app [9,10]. Simulations estimate that for the app to contain the epidemic, at least 56% of a country's population needs to use the app and comply with the app's recommendations [11], although lower uptake rates are also effective in reducing the number of infections [12]. This paper examines to what extent this goal is likely to be achieved in Germany by providing answers to the following research question: What proportions of the general population aged 18 to 77 years in Germany (1) have access to the devices required to install the official contact tracing app, (2) are able to install it, and (3) are willing to install the app, use it, and act according to its recommendations?

Our predictions show that the adoption rate of 56% needed to contain the epidemic will be missed by a considerable margin. However, contact tracing apps may still be effective if specific subgroups adopted them at a higher rate. In particular, if a high proportion of persons who are frequently in contact with persons outside their household (ie, potential spreaders) adopted the app, its spread may be significantly curbed. In a similar vein, if a high proportion of persons who are likely to get severely ill from the disease (ie, vulnerable groups) adopted the app, health workers may be able to treat them early on and, thus, decrease the impact of COVID-19. Therefore, we investigate adoption rates among these two population subgroups by asking the following research questions:

1. What proportions of potential spreaders (1) have access to the devices required to install the official contact tracing app, (2) are able to install it, and (3) are willing to install the app, use it, and act according to its recommendations?
2. What proportions of persons with high vulnerability to a serious infection (1) have access to the devices required to install the official contact tracing app, (2) are able to install it, and (3) are willing to install the app, use it, and act according to its recommendations?

The official COVID-19 contact tracing app in Germany, the Corona-Warn-App, can be downloaded from the Apple App Store or Google Play free of charge and installed on iPhones, with iOS version 13.5 or higher, and Android smartphones, with Android version 6.0 or higher [9,10]. The app can be installed by the same person on multiple devices. Once installed, the app detects other app users in proximity by exchanging encrypted ID numbers between devices using Bluetooth Low Energy technology. The ID numbers change constantly and are stored locally on the device, relying on a decentralized approach for data storage. The user's geolocation is not tracked. The app automatically informs users when they have been in contact with someone confirmed infected with SARS-CoV-2 and provides behavioral recommendations, including domestic quarantine and tests for SARS-CoV-2. The identity of the person using the app remains anonymous. An app user with a positive test result can enter this result into the Corona-Warn-App. By doing so, all proximity contacts are automatically notified of their own potential infection. Users can deactivate and reactivate the COVID-19 exposure notifications at any time or can completely uninstall the app. Using the Corona-Warn-App is voluntary and meets the European Union General Data Protection Regulation [13,14].

There are several potential barriers that may prevent people from using an app [15,16]. An initial barrier is access to a smartphone capable of installing the desired app and access to the internet [17,18]. In the case of the Corona-Warn-App, persons additionally need a smartphone with an iOS or Android operating system [9,10]. Among smartphone users with compatible devices, a second barrier is their ability to carry out the tasks required to operate the app [19]. The Corona-Warn-App requires the user to have the ability to download and install the app and to handle Bluetooth [9,10]. A final potential barrier is a person's willingness to use the contact tracing app. A key correlate of this barrier in Germany seems to be privacy concerns regarding the sharing of personal data and distrust in unfamiliar technology and processes running in the background [20,21].

The effectiveness of contact tracing apps not only hinges on access, ability, and willingness to use such an app, but also on *how* individuals use the app. People need to carry their smartphone with them throughout the day, regularly recharge the smartphone batteries, keep their smartphone turned on, and keep the contact tracing feature activated so that the app can detect proximity contacts at all times. For some activities, however, people usually do not take their phone with them; for example, while exercising. As a result, the contacts during these periods are not being tracked. In addition, app-based contact tracing is subject to technical limitations, such as Bluetooth-based measurement errors, which may cause errors in the contacts detected [22].

COVID-19 exposure apps have been developed in many countries [23-25]. The MIT Technology Review's Covid Tracing Tracker currently lists 47 countries with available or soon-to-be available contact tracing apps [23], yet installation rates across countries are low. For example, there were only 22.4 million app downloads in Germany as of November 12, 2020, around 5 months after its introduction, compared to a

population of 83.2 million [10], even though an early Oxford-led study suggested high support for contact tracing apps of 74.8% across five countries: France, Germany, Italy, the United Kingdom, and the United States [26]. However, the Oxford study comes with a major caveat: the predictions were based on nonprobability online samples, which are known to be self-selective and severely overrepresent technologically interested persons; thus, they do not accurately represent likely behaviors in those countries' populations [27]. The selective nature of the data may, therefore, explain the discrepancy between the high support for the apps in the Oxford study and observed installation rates.

Methods

Data

To allow for timely and accurate population predictions of the adoption of the Corona-Warn-App, we based our analyses on data collected close to the launch of the app and on a probability sample of the general population aged 18 to 77 years. In this section, we describe key aspects of our data collection according to the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [28].

The data were collected in the Mannheim Corona Study (MCS) [29]. The MCS was implemented within the German Internet Panel (GIP), a long-standing, offline-recruited, probability-based online panel of the general adult population in Germany. GIP sample members were recruited in 2012, 2014, and 2018. The 2012 and 2014 recruitments were based on area probability samples with full address listings and face-to-face recruitment interviews [30]. Persons in households without internet and/or computer access were provided with user-friendly devices, internet connections, and/or information technology support to enable their participation in the panel [31]. The 2018 recruitment was based on a probability sample drawn from municipal population registers and initial postal invitations. Sample members were informed about the scope of the study, the investigator, as well as how their data would be stored, and they consented to their participation and data storage.

A subsample of the GIP was invited to participate in the MCS that was conducted for 16 weeks from March 20 to July 10, 2020. The MCS was fielded with a rotating daily panel design. Each week the same MCS sample members were invited via email to participate in the study on the same day of the week and were re-interviewed on a variety of social, psychological, and economic topics [29]. They were sent a personalized link to the survey or were able to log in to the study website using their username and password to access the survey. Upon survey completion, respondents received an incentive of €2 (US $2.40). Our study on the Corona-Warn-App, which was launched on June 16, 2020, was implemented within the MCS in week 13 (ie, June 12 to 19, 2020). The questionnaire from that week contained 44 pages, with one item per page. Respondents were able to change their answers using a *back* button on most pages. Prior to fielding the study, the usability and technical functionality of the questionnaire was tested. A total of 5427 persons, aged 18 to 77 years, were invited to participate in our study, of which 3276 responded (60.4%).

To correct for a potential overrepresentation of persons with higher digital affinity [32], all predictions were weighted with a two-stage weighting procedure. Our weighting accounted for potential coverage, sampling, and nonresponse biases of the online data collection [33,34]. At the first stage, we estimated a response propensity weight, which projected the characteristics of the MCS respondents to the GIP recruitment samples. The weighting characteristics for the 2012 and 2014 samples included computer and internet access within the household; weighting characteristics for the 2018 sample included frequency of internet use, intensity of internet use, computer use, smartphone use, tablet use, and importance of up-to-date technology. At the second stage, we estimated a raking weight, which extrapolated the characteristics of the MCS respondents to the general population according to the Mikrozensus, that is, official statistics provided by the German Federal Statistical Office [35]. The weighting characteristics included age, gender, marital status, highest level of education, household size, and federal state. Missing values on the weighting variables were imputed with a chained-equations algorithm [36]. The final weight was trimmed for values greater than 4 and values less than 0.25. Despite the weighting procedure, our analyses were still likely to overestimate the app adoption rate in the general population, which we further address in the Discussion section.

Measures

Access, Ability, and Willingness

We measured adoption rates and potential barriers to adoption in sequential sets of survey questions and estimated (1) the population's access to and use of compatible smartphones, (2) their ability to install and correctly use the app, and (3) their willingness to adopt the app and act according to its instructions.

Through three questions, we estimated people's access: "Do you personally use a smartphone?"; if *yes*, "Which of the following types best describes your smartphone?" and "How often do you carry your smartphone with you when you leave the house?" (see Multimedia Appendix 1). We defined persons as having access to the app if they own an iPhone or Android phone and carry it with them at least most of the time when they leave the house. We did not differentiate between operating system versions and were, thus, likely to overestimate access to the app. However, research about the distribution of operating system versions installed on smartphones in Germany suggests that the large majority of Android smartphones and iPhones have the version installed that is required for the Corona-Warn-App to work [37,38].

People's ability to use the app was measured through four questions: "Do you know how to install an app, i.e. an additional program, on your smartphone?"; if *no* or *not sure*, "Do you know anyone who could help you with installing the Corona-Warn-App on your smartphone, e.g. family, friends, or neighbors?"; "Do you know how to activate Bluetooth on your smartphone?"; and if *no* or *not sure*, "Do you know anyone who could help you with the activation of Bluetooth on your smartphone, e.g. family, friends, or neighbors?" (see Multimedia Appendix 1). A limitation of the two questions about the potential help of family, friends, or neighbors is that they do not differentiate between the usual situation before the

COVID-19 lockdown measures came into effect and the present situation: while some individuals may generally know plenty of people who can assist them with technology-related issues, they may not be able to meet these people due to the lockdown measures. However, since infection rates fell in June 2020 and the lockdown measures were gradually lifted, this limitation likely ceased to affect adoption rates during the summer. We defined persons as able to use the app if they know how to install an app or have someone who can help them with it, and if they know how to activate Bluetooth on their smartphone or have someone who can help them with it. Since access is a necessary condition for being able to install the Corona-Warn-App on a smartphone, persons who were defined as not having access were also defined as not being able to use the app.

Finally, we measured people's willingness to correctly use the app through four questions: "Would you install the official Corona-Warn-App on your smartphone when it is available?"; if at least *probably not install*, "Would you follow the request of the Corona-Warn-App and go into domestic quarantine as a precaution?"; "Would you comply with the request of the Corona-Warn-App and get tested for the virus?"; and "Would you enter the test result into the Corona-Warn-App if you were tested positive for the virus?" (see Multimedia Appendix 1). We defined persons as willing to correctly use the app if they are probably or definitely willing to install the app, if they are probably or definitely willing to quarantine if requested, if they are probably or definitely willing to get tested if requested, and if they are probably or definitely willing to enter their own test result into the app if they were tested positive. Since access and ability are necessary conditions for being willing to install the Corona-Warn-App on their smartphone, persons who were defined as not having access or not being able to use the app were also defined as not being willing to correctly use the app.

Potential to Spread SARS-CoV-2 and Potential to Be at Risk of COVID-19

Two variables from the MCS and GIP data collection classified persons according to their potential for spreading the virus: number of social contacts within the past 7 days and employment situation, both collected from the MCS in week 13 (see Multimedia Appendix 1). The resulting variable has the following categories:

1. Met socially with other persons several times in the past 7 days and worked full time outside the home.
2. Met socially with other persons several times in the past 7 days but did not work full time or did not work outside the home.
3. Met socially with other persons once or less often in the past 7 days but worked full time outside the home.
4. Met socially with other persons once or less often in the past 7 days and did not work full time or did not work outside the home.

In addition, two variables from the MCS and GIP data collection classified persons according to their potential for being vulnerable to a serious infection: being aged 60 to 77 years, collected from the GIP, and having any health condition that, according to the RKI, may be correlated with an increased risk, collected from the MCS in week 13 (see Multimedia Appendix 1). The resulting variable has the following categories:

1. Aged 60 to 77 years and with at-risk health conditions.
2. Aged 60 to 77 years but without at-risk health conditions.
3. Aged 18 to 59 years but with at-risk health conditions.
4. Aged 18 to 59 years and without at-risk health conditions.

Although participants were not required to respond to all questions in the MCS survey, the amount of missing data was low for frequency of social contacts (8 missing values), work outside home (1 missing value), and health condition (5 missing values).

Analytical Strategy

First, we reported overall rates of Corona-Warn-App adoption, distinguishing the three levels of potential barriers: access, ability, and willingness. Subsequently, we estimated separate adoption rates by the potential to spread SARS-CoV-2 and the potential to be vulnerable to COVID-19. All estimations were weighted as described above to enable reliable population predictions. Adoption rates across subgroups were reported by means of the predicted probabilities of a logistic regression, not including any covariates. Using the margins command in Stata 16.0 (StataCorp LLC), predicted probabilities were computed to conduct chi-square tests of differences in adoption rates across subgroups. Finally, we examined whether the introduction of the Corona-Warn-App during our data collection period influenced people's willingness to install and use the app. For this purpose, we estimated a logistic regression for willingness on a dummy variable identifying whether our data were collected before or after the publication of the Corona-Warn-App, controlling for key sociodemographic characteristics (see Multimedia Appendix 1).

Results

For the overall rate of adoption of the Corona-Warn-App, we estimated that 37.9% of the population in Germany aged 18 to 77 years have access to, are able to, and are willing to install the app (see Figure 1 and Multimedia Appendix 2). Asked whether they would be willing to go into domestic quarantine and get tested when requested to do so by the app, these rates reduce to 34.9% and 37.3%, respectively. If tested positive, 37.6% of the population aged 18 to 77 years would be willing to enter the test result into the app.

Whereas a lack of willingness is the foremost barrier to app adoption, access also plays a considerable role. Only 91.8% of the population aged 18 to 77 years uses a smartphone, 88.5% uses one with a compatible operating system, and 85.0% carries it with them most or all of the time when outside the house. An inability to install apps and handle Bluetooth further reduces potential adoption rates to 81.3% and 81.8%, respectively.

Next, we examined whether higher adoption rates were achieved among the relevant subgroups of potential spreaders (see Figure 2 and Multimedia Appendix 3) and the potentially vulnerable (see Figure 3 and Multimedia Appendix 4).

Figure 1. Predicted adoption rates by access, ability, and willingness (N=3276). Error bars represent 95% CI.

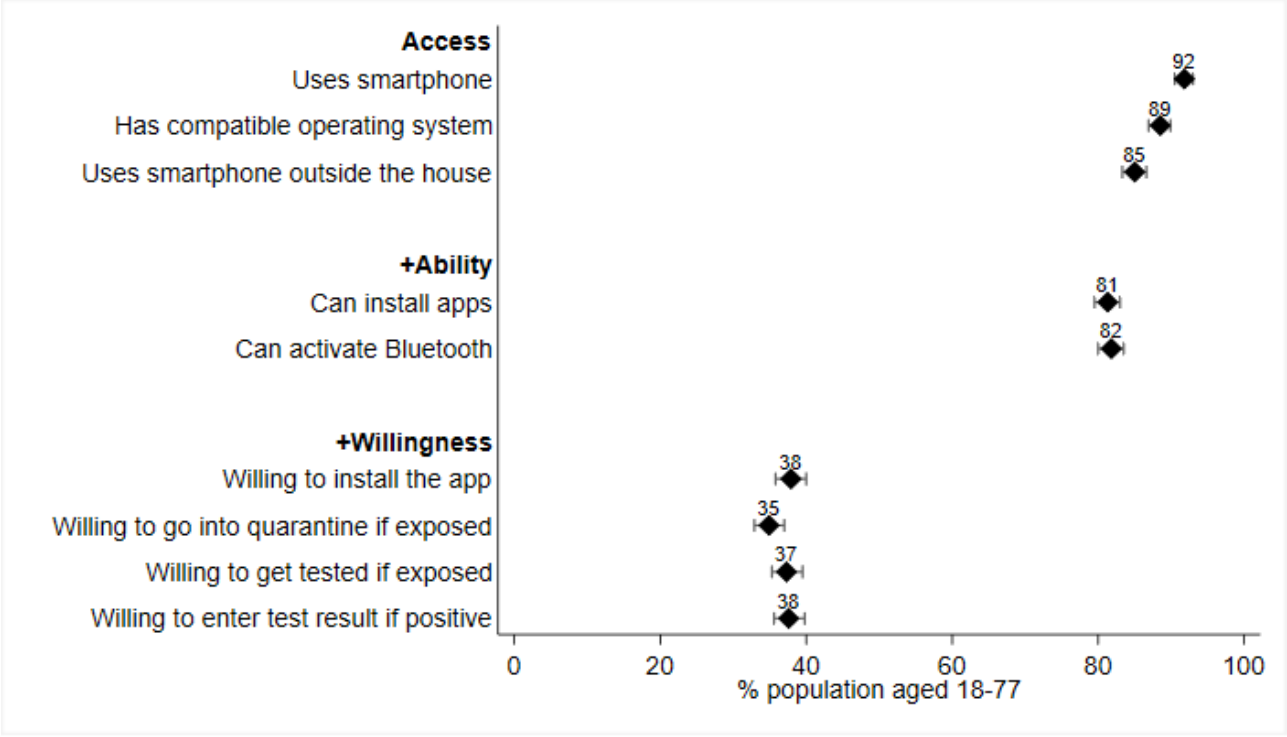
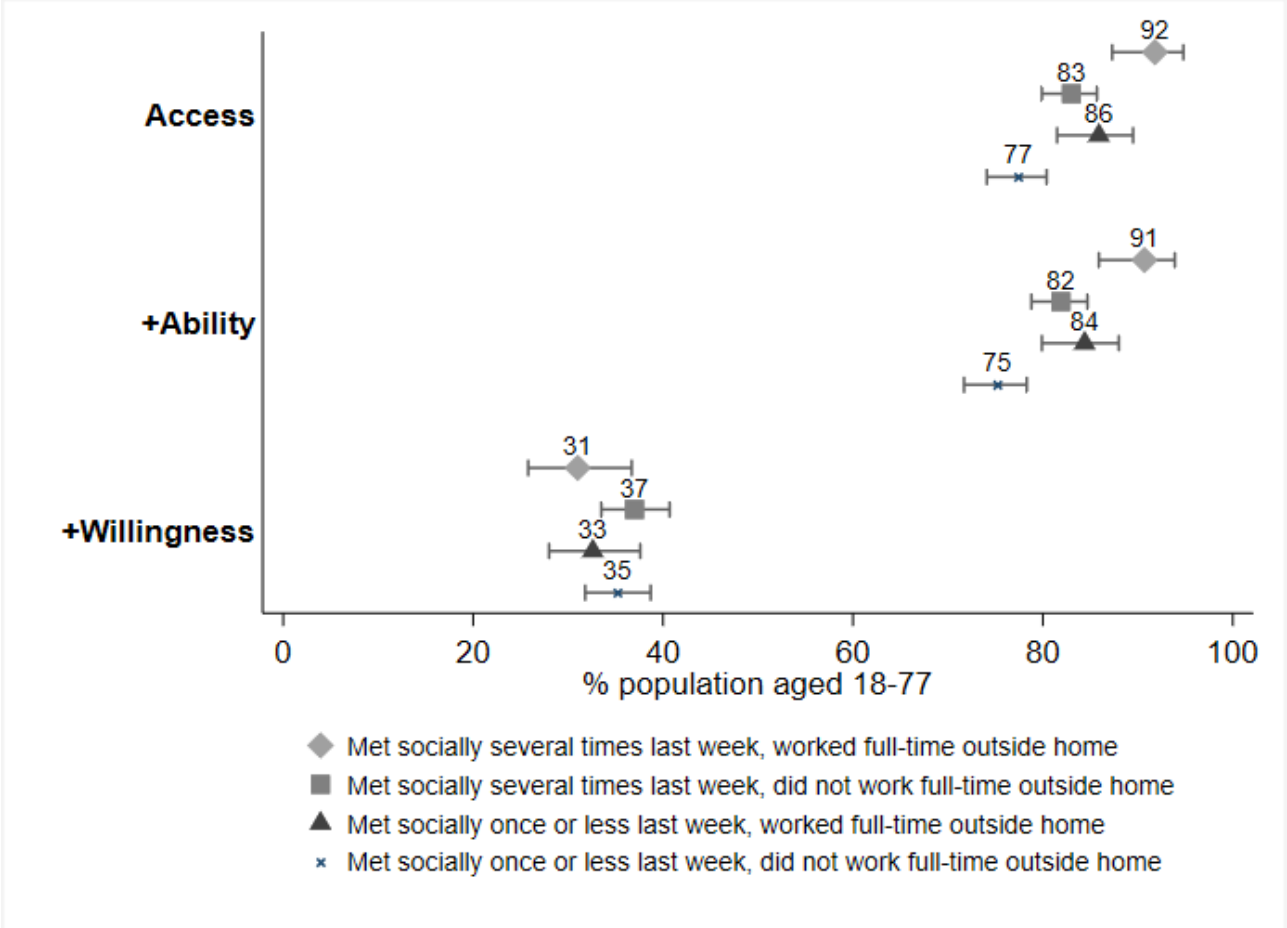


Figure 2. Predicted adoption rates by potential to spread SARS-CoV-2 (access: N=3267; ability: N=3267; willingness: N=3266). Error bars represent 95% CI.

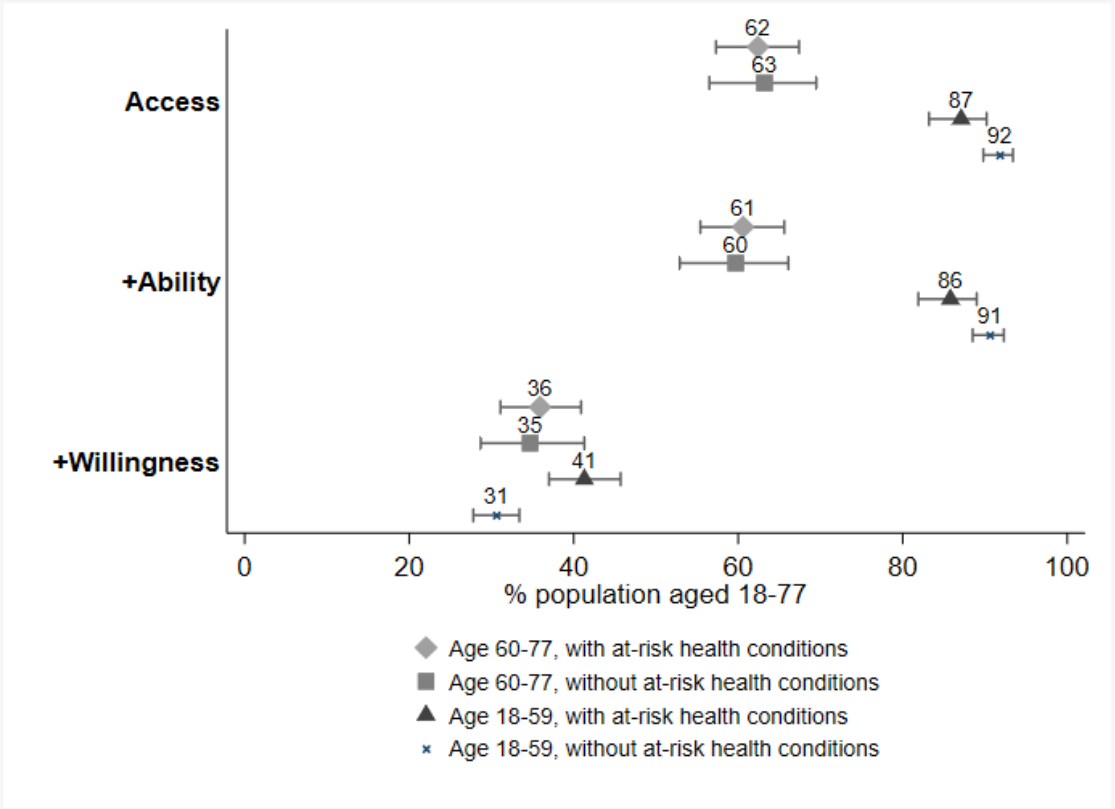


Persons with a high potential to spread the virus (ie, met socially several times last week and worked full time outside the home: 91.8%) are significantly more likely to have access than those with a medium potential to spread the virus (ie, met socially several times last week and did not work full time outside the home: 83.0%; and met socially once or less often last week and worked full time outside the home: 85.9%). Persons with a medium potential to spread the virus are, in turn, significantly more likely to have access than persons with a low potential to spread the virus (ie, met socially once or less often last week and did not work full time outside the home: 77.4%). The same pattern of significant group differences was found for ability (90.7% vs 81.9% and 84.4% vs 75.2%). However, in predicting the overall adoption rates (ie, access + ability + willingness), we did not find any significant group differences (ie, met socially several times last week and worked full time outside the home: 31.0%; met socially several times last week and did not work full time outside the home: 37.0%; met socially once or less last week and worked full time outside the home: 32.6%; and met socially once or less last week and did not work full time outside the home: 35.2%). In fact, the pattern does not deliver support for the hope that the contact tracing app may be more effective in the subgroup with a high potential to spread the virus, with similarly low overall adoption rates for those with a high potential to spread SARS-CoV-2 compared to those with a medium or low potential.

When examining the characteristics of those with a high potential to spread SARS-CoV-2 but unwilling to install and correctly use the Corona-Warn-App, we found that the large majority (68%) are between the ages of 30 and 59 years, with an additional 22% between 18 and 29 years and 10% between 60 and 77 years. Furthermore, most of these individuals are male (67%), with an intermediate (42%) or higher education (40%) as opposed to a lower education (18%). Only 10% feel personally threatened by COVID-19, while a majority (65%) think that the economic damage of the measures taken by governments to fight the pandemic is greater than their benefit for society. Interestingly, privacy concerns do not seem to be the driving factors that influence their decision to not adopt the app, since only 13% indicated they are very concerned about their privacy.

Regarding potential vulnerability, we observed an age effect on access and ability (see Figure 3 and Multimedia Appendix 4). The older age groups (ie, aged 60-77 years, with at-risk health conditions; and aged 60-77 years, without at-risk health conditions) are significantly less likely than younger age groups (ie, aged 18-59 years, with at-risk health conditions; and aged 18-59 years, without at-risk health conditions) to use a compatible smartphone (62.4% and 63.2% vs 87.1% and 91.8%) and to be able to install and use the app (60.6% and 59.7% vs 85.8% and 90.6%), with very large differences independent of at-risk health conditions. In predicting the overall adoption rates (ie, access + ability + willingness), we did not find any consistent significant differences across vulnerability groups (ie, aged 60-77 years, with at-risk health conditions: 35.9%; aged 60-77 years, without at-risk health conditions: 34.7%; aged 18-59 years, with at-risk health conditions: 41.3%; and aged 18-59 years, without at-risk health conditions: 30.6%).

Figure 3. Predicted adoption rates by potential vulnerability to COVID-19 (access: N=3270; ability: N=3270; willingness: N=3269). Error bars represent 95% CI.



When examining the characteristics of those with a high vulnerability to COVID-19 who do not have access and are unable to use the app, we found that the majority (43%) are older than 70 years, with an additional 28% between 60 and 64 years and 29% between 65 and 69 years. Most of these individuals are living in lower-income households, with a monthly net income between €0 (US $0) and €1999 (US $2414) (41%) or between €2000 (US $2415) and €2999 (US $3623) (39%), as opposed to those living in higher-income households (ie, between €3000 [US $3624] and €3999 [US $4830]: 11%; and €4000+ [US $4831+]: 10%).

Finally, when examining whether persons who were interviewed before the introduction of the Corona-Warn-App showed different adoption rates than persons interviewed after the app launch, we found no significant differences (see Multimedia Appendix 5).

Discussion

The official contact tracing app by the German federal government and the center for disease control and prevention, RKI, was introduced on June 16, 2020. The Corona-Warn-App was heavily advertised by government officials and health representatives as an effective way to contain the spread of SARS-CoV-2. According to epidemiological models, however, 56% of the population needs to adopt the app for it to contain the epidemic [11].

Our study shows that the 56% target mark will likely be missed by a considerable margin. For the population aged 18 to 77 years, our estimations predict an overall adoption rate of 34.7%. The largest barrier is people's willingness to install and correctly use the app; however, access to a compatible smartphone and the ability to install the app also play roles. Given the age groups covered in our study, we consider this an optimistic estimate. For cohorts aged 78 years and over and children, the adoption rates are likely considerably lower.

Persons with the highest potential to spread the virus (ie, with frequent social and work contacts) are more likely to have access and the ability to use the app (90.7%) than the average in the population aged 18 to 77 years (81.0%). Overall, persons with a high potential to spread the virus are no more likely to adopt the app than persons with fewer social and work interactions.

Persons at risk to fall seriously ill or die from an infection (ie, those aged 60 to 77 years with at-risk health conditions) have significantly reduced access and ability to use the app (60.6%) compared to the average in the population aged 18 to 77 years (81.0%). Those who can use the app, however, are overwhelmingly willing to do so. As a consequence, persons with high vulnerability to COVID-19 are equally likely to adopt the app as are less vulnerable groups.

Overall, the findings imply a pessimistic view on the effectiveness of app-based contact tracing to contain the COVID-19 pandemic in Germany, with low adoption rates in the general population and issues of selectivity across subgroups as noted by Klingwort and Schnell [22]. In addition to low uptake in the general population, vulnerable groups who would benefit from an efficient contact tracing approach have limited smartphone coverage and limited ability to use the app. Furthermore, those with a high potential to spread SARS-CoV-2 who would have the necessary devices and abilities to install the app are predominantly unwilling to do so. Even though, as Hinch et al [11] pointed out, uptake rates of contact tracing apps lower than the 56% target may still contribute to a reduction in the number of infections, Germany will miss the 56% target by a huge margin and would do well investing in additional routes of tracing potentially infected individuals.

Our study was conducted during the week the Corona-Warn-App was introduced in Germany. This enabled us to implement a questionnaire that considers all technological and data privacy specifications of the actual app. The findings also allow us to formulate actionable policy recommendations. First, we recommend targeting information campaigns at groups with a high potential to spread the virus but who are unwilling to install and correctly use the Corona-Warn-App, in particular men and those aged between 30 and 59 years, to encourage them to adopt the app. Our second recommendation is to invest further resources to provide vulnerable groups of the population, in particular older individuals and those in lower-income households, with the necessary devices and assistance to overcome their specific barriers to app adoption.

This study is not free from limitations. First, the data were collected from an online panel. Although individuals without computer or internet access were provided with the necessary equipment and support, and weights were used in all analyses to correct for coverage and nonresponse biases, we cannot rule out that the data still overrepresent individuals with an interest in technology. When invited to the MCS, panel members had already completed online surveys over the course of at least 2 years. They are, thus, more likely to be interested in digital technologies, such as the Corona-Warn-App, than their counterparts who dropped out of the online panel.

Second, we can expect panel members who agreed to participate in the MCS and be interviewed every week on topics related to the COVID-19 pandemic to be more interested in contributing to a better understanding of the social impacts of the pandemic and possibly be more concerned than the average citizen. Such traits may also affect their willingness to install a contact tracing app.

Third, our analyses are limited to individuals aged between 18 and 77 years. We are, thus, missing sizable population groups: those aged 0 to 10 years make up 10% of the general population in Germany, those aged 11 to 17 years constitute 6%, and those aged 78 years or older make up 9% [39]. The youngest age group might be disregarded in an estimation of the effectiveness of the Corona-Warn-App since they predominantly move within small, defined social circles and are unlikely to carry smartphones with them at all times. The age group of 11 to 17 years is likely to have a high potential to spread SARS-CoV-2 but possibly low interest in adopting the Corona-Warn-App, whereas the oldest age group is highly vulnerable to COVID-19 and is also likely to have low app adoption rates because of limited smartphone access. As a result of all of these limitations, our predicted adoption rates are an optimistic view of the

situation. True values in the general population are likely to paint an even more pessimistic reality.

Fourth, our study is based on reported hypothetical behavior rather than actual behavior. Although hypothetical measures of willingness to install an app are subjective and may be subject to various response errors, such as social desirability or recall errors [40], these measures were shown to be correlated with actual behavior in previous studies [15,16,20,41]. In weeks 14 to 16 of the MCS, fielded between June 19 and July 10, 2020, we also collected data about whether people installed the Corona-Warn-App on their smartphone. The results suggest that the actual installation rate is very similar to the app adoption rate estimated in this paper: by July 10, 2020, almost 1 month after the app was introduced, 36% (95% CI 34%-38%) of the population between 18 and 77 years had installed the app, 55% (95% CI 53%-58%) had not installed the app, 1% (95% CI 1%-2%) had installed the app but had uninstalled it since then, and 7% (95% CI 6%-8%) did not use a smartphone. We also compared the responses about hypothetical and actual willingness to install the app among 2877 survey respondents who completed the questions both in week 13 and week 16 of the MCS survey, combining the categories *app not installed*; *app installed, but uninstalled since then*; and *don't use a smartphone*. The results show a correlation of r=0.6 (P<.001), which gives us confidence that the hypothetical measures used in this paper are rather accurate.

A potential avenue of future research would be to study whether the use of such a contact tracing app changes people's behavior. After the installation, individuals' behavior may become riskier (eg, less compliant toward social distancing measures), since the app may give them a feeling of security. If such a behavior is prevalent in the population, this may further reduce the effectiveness of app-based contact tracing.

Regarding data availability, the GIP data used in the analyses of this article are freely available as part of the GIP Scientific Use Files. They can be requested from the GESIS Data Archive for the Social Sciences (GESIS-DAS) [42]. The MCS data are envisioned to be published as Scientific Use Files by the end of 2021 at the latest. Until then, these data can be accessed at the Onsite Data Access facilities of the GIP Secure Data Center located at the Collaborative Research Center *Political Economy of Reforms* (SFB 884), University of Mannheim, B6 30-32, Mannheim, Germany. Researchers wishing to make use of the Onsite Data Access facilities may contact secretary@reforms.uni-mannheim.de. Researchers wishing to get access to the analysis code may contact the corresponding author.

Acknowledgments

We thank Elena Madiai, Lisa Wellinghoff, Katja Sonntag, and Sabrina Seidl for excellent research assistance. The GIP and the MCS are funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) through the Collaborative Research Center *Political Economy of Reforms* (SFB 884; project ID: 139943784). The data collection for the MCS is also funded by the German Federal Ministry for Work and Social Affairs (BMAS; project ID: FIS.00.00185.20).

Authors' Contributions

AB was responsible for the study conceptualization and methodology, funding acquisition, study investigation, supervision, obtaining resources, writing the original draft, and reviewing and editing subsequent drafts. AW was responsible for the study conceptualization and methodology, data analysis, study investigation, data visualization, writing the original draft, and reviewing and editing subsequent drafts. CC was responsible for the study investigation and methodology. TR was responsible for data curation, study investigation, and project administration. MF was responsible for study investigation and project administration. SF was responsible for data curation and study investigation. KM, EN, and MR were responsible for study investigation. UK was responsible for study investigation and methodology as well as supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study questionnaire.
[PDF File (Adobe PDF File), 182 KB - jmir_v23i3e23362_app1.pdf]

Multimedia Appendix 2
Predicted adoption rates of the COVID-19 contact tracing app in Germany by access, ability, and willingness.
[PDF File (Adobe PDF File), 80 KB - jmir_v23i3e23362_app2.pdf]

Multimedia Appendix 3
Predicted adoption rates of the COVID-19 contact tracing app in Germany by potential to spread SARS-CoV-2.
[PDF File (Adobe PDF File), 82 KB - jmir_v23i3e23362_app3.pdf]

Multimedia Appendix 4
Predicted adoption rates of the COVID-19 contact tracing app in Germany by potential vulnerability to COVID-19.
[PDF File (Adobe PDF File), 79 KB - jmir_v23i3e23362_app4.pdf]

Multimedia Appendix 5
Results of a logistic regression of willingness to use the COVID-19 contact tracing app.
[PDF File (Adobe PDF File), 142 KB - jmir_v23i3e23362_app5.pdf]

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19. World Health Organization. 2020 Mar 11. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-06-24]
2. Hale T, Angrist N, Cameron-Blake E, Hallas L, Kira B, Majumdar S, et al. Oxford COVID-19 Government Response Tracker. Blavatnik School of Government. 2020. URL: https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker [accessed 2020-06-26]
3. Askitas N, Tatsiramos K, Verheyden B. Lockdown Strategies, Mobility Patterns and COVID-19. Discussion Paper. Bonn, Germany: IZA Institute of Labor Economics; 2020 May. URL: http://ftp.iza.org/dp13293.pdf [accessed 2020-06-26]
4. Deb P, Furceri D, Ostry J, Tawk N. IMF Working Paper: The Economic Effects of COVID-19 Containment Measures. Washington, DC: International Monetary Fund; 2020 Aug 07. URL: https://www.imf.org/en/Publications/WP/Issues/2020/08/07/The-Economic-Effects-of-COVID-19-Containment-Measures-49571 [accessed 2021-01-22]
5. Adams-Prassl A, Boneva T, Golin M, Rauh C. The Impact of the Coronavirus Lockdown on Mental Health: Evidence from the US. HCEO Working Paper. Chicago, IL: Human Capital and Economic Opportunity Global Working Group; 2020 May. URL: http://humcap.uchicago.edu/RePEc/hka/wpaper/Adams-Prassl_Boneva_Golin_etal_2020_impact-coronavirus-mental-health.pdf [accessed 2021-02-23]
6. Kontaktpersonen-Nachverfolgung bei SARS-CoV-2-Infektionen. Robert Koch Institute. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Kontaktperson/Management.html [accessed 2020-06-26]
7. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):eabb6936 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
8. Böhmer MM, Buchholz U, Corman VM, Hoch M, Katz K, Marosevic DV, et al. Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: A case series. Lancet Infect Dis 2020 Aug;20(8):920-928 [FREE Full text] [doi: 10.1016/S1473-3099(20)30314-5] [Medline: 32422201]
9. Coronavirus warning app (Corona-Warn-App). Federal Government of Germany. URL: https://www.bundesregierung.de/breg-de/themen/corona-warn-app/corona-warn-app-englisch [accessed 2020-06-26]
10. Infektionsketten digital unterbrechen mit der Corona-Warn-App. Robert Koch Institute. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Warn_App.html [accessed 2020-11-17]
11. Hinch R, Probert W, Nurtay A, Kendall M, Wymant C, Hall M, et al. Effective configurations of a digital contact tracing app: A report to NHSX. The Conversation. 2020 Apr 16. URL: https://cdn.theconversation.com/static_files/files/1009/Report_-_Effective_App_Configurations.pdf?1587531217 [accessed 2021-02-23]
12. Abueg M, Hinch R, Wu N, Liu L, Probert W, Wu A, et al. Modeling the combined effect of digital exposure notification and non-pharmaceutical interventions on the COVID-19 epidemic in Washington state. medRxiv. Preprint posted online on September 2, 2020. URL: https://www.medrxiv.org/content/10.1101/2020.08.29.20184135v1 [accessed 2020-09-09]
13. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). EUR-Lex. 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj [accessed 2021-02-18]
14. Privacy notice. Corona-Warn-App. URL: https://www.coronawarn.app/assets/documents/cwa-privacy-notice-en.pdf [accessed 2020-06-26]
15. Jäckle A, Burton J, Couper MP, Lessof C. Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: Coverage and participation rates and biases. Surv Res Methods 2019 Apr 11;13(1):23-44. [doi: 10.18148/srm/2019.v1i1.7297]
16. Wenz A, Jäckle A, Couper MP. Willingness to use mobile technologies for data collection in a probability household panel. Surv Res Methods 2019 Apr 11;13(1):1-22. [doi: 10.18148/srm/2019.v1i1.7298]
17. Keusch F, Bähr S, Haas G, Kreuter F, Trappmann M. Coverage error in data collection combining mobile surveys with passive measurement using apps: Data from a German national survey. Sociol Methods Res 2020 Apr 07:1-38 [FREE Full text] [doi: 10.1177/0049124120914924]
18. Antoun C, Conrad FG, Couper MP, West BT. Simultaneous estimation of multiple sources of error in a smartphone-based survey. J Surv Stat Methodol 2019 Mar;7(1):93-117. [doi: 10.1093/jssam/smy002]

19. Hargittai E. Second-level digital divide: Differences in people's online skills. First Monday 2002 Apr 01;7(4). [doi: 10.5210/fm.v7i4.942]
20. Kreuter F, Haas G, Keusch F, Bähr S, Trappmann M. Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. Soc Sci Comput Rev 2020 Oct 01;38(5):533-549. [doi: 10.1177/0894439318816389]
21. Felderer B, Blom AG. Acceptance of the automated online collection of geographical information. Sociol Methods Res 2019 Dec 05:1-21 [FREE Full text] [doi: 10.1177/0049124119882480]
22. Klingwort J, Schnell R. Critical limitations of digital epidemiology: Why COVID-19 apps are useless. Surv Res Methods 2020 Jun 02;14(2):95-101. [doi: 10.18148/srm/2020.v14i2.7726]
23. Howell O'Neill P, Ryan-Mosley T, Johnson B. A flood of coronavirus apps are tracking us. Now it's time to keep track of them. MIT Technology Review. 2020 May 07. URL: https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/ [accessed 2020-11-17]
24. Singh HJL, Couch D, Yap K. Mobile health apps that help with COVID-19 management: Scoping review. JMIR Nurs 2020;3(1):e20596 [FREE Full text] [doi: 10.2196/20596] [Medline: 32897271]
25. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: Challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
26. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: Cross-country survey study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
27. Cornesse C, Blom AG, Dutwin D, Krosnick JA, de Leeuw ED, Legleye S, et al. A review of conceptual approaches and empirical evidence on probability and nonprobability sample survey research. J Surv Stat Methodol 2020 Feb 01;8(1):4-36. [doi: 10.1093/jssam/smz041]
28. Eysenbach G. Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
29. Blom AG, Cornesse C, Friedel S, Krieger U, Fikel M, Rettig T, et al. High-frequency and high-quality survey data collection: The Mannheim Corona Study. Surv Res Methods 2020 Jun 02;14(2):171-178. [doi: 10.18148/srm/2020.v14i2.7735]
30. Blom AG, Gathmann C, Krieger U. Setting up an online panel representative of the general population. Field methods 2015 Mar 26;27(4):391-408. [doi: 10.1177/1525822x15574494]
31. Blom AG, Herzing JME, Cornesse C, Sakshaug JW, Krieger U, Bossert D. Does the recruitment of offline households increase the sample representativeness of probability-based online panels? Evidence from the German Internet Panel. Soc Sci Comput Rev 2017 Aug 01;35(4):498-520. [doi: 10.1177/0894439316651584]
32. Herzing JME, Blom AG. The influence of a person's digital affinity on unit nonresponse and attrition in an online panel. Soc Sci Comput Rev 2019 Jun 01;37(3):404-424. [doi: 10.1177/0894439318774758]
33. Little RJA, Rubin DB. Statistical Analysis with Missing Data. Second Edition. Hoboken, NJ: John Wiley & Sons; 2002.
34. Little RJA. Survey nonresponse adjustments for estimates of means. Int Stat Rev 1986 Aug;54(2):139-157. [doi: 10.2307/1403140]
35. Microcensus. Research Data Centres of the Statistical Offices of the Federation and the Federal States of Germany. URL: https://www.forschungsdatenzentrum.de/en/household/microcensus [accessed 2020-06-26]
36. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw 2011 Dec;45(3):1-67 [FREE Full text] [doi: 10.18637/jss.v045.i03]
37. StatCounter. URL: https://gs.statcounter.com/android-version-market-share/mobile-tablet/germany [accessed 2020-09-09]
38. StatCounter. URL: https://gs.statcounter.com/ios-version-market-share/mobile-tablet/germany [accessed 2020-09-09]
39. Bevölkerung: Deutschland, Stichtag, Altersjahre [Population: Germany, record date, years of age]. Database of the Federal Statistical Office of Germany. URL: https://www-genesis.destatis.de/genesis/online/data [accessed 2020-09-09]
40. Sudman S, Bradburn NM, Schwarz N. Thinking About Answers: The Application of Cognitive Processes to Survey Methodology. San Francisco, CA: Jossey-Bass; 1996.
41. Keusch F, Struminskaya B, Antoun C, Couper MP, Kreuter F. Willingness to participate in passive mobile data collection. Public Opin Q 2019 Jul;83(Suppl 1):210-235 [FREE Full text] [doi: 10.1093/poq/nfz007] [Medline: 31337924]
42. GN0109: German Internet Panel. GESIS Leibniz Institute for the Social Sciences. URL: https://dbk.gesis.org/dbksearch/GDesc2.asp?no=0109&tab=&ll=10&notabs=1&db=E [accessed 2021-02-18]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
GESIS-DAS: GESIS Data Archive for the Social Sciences
GIP: German Internet Panel
MCS: Mannheim Corona Study
RKI: Robert Koch Institute

Edited by G Eysenbach, R Kukafka; submitted 10.08.20; peer-reviewed by U Kohler, M Revilla; comments to author 15.08.20; revised version received 09.09.20; accepted 18.01.21; published 02.03.21.

Please cite as:
Blom AG, Wenz A, Cornesse C, Rettig T, Fikel M, Friedel S, Möhring K, Naumann E, Reifenscheid M, Krieger U
Barriers to the Large-Scale Adoption of a COVID-19 Contact Tracing App in Germany: Survey Study
J Med Internet Res 2021;23(3):e23362
URL: https://www.jmir.org/2021/3/e23362
doi:10.2196/23362
PMID:33577466

Original Paper

Social Media Engagement and Influenza Vaccination During the COVID-19 Pandemic: Cross-sectional Survey Study

Arriel Benis[1,2], PhD; Anna Khodos[1], BSc; Sivan Ran[1], BSc; Eugene Levner[3], PhD; Shai Ashkenazi[4], MD, MSc

[1]Faculty of Industrial Engineering and Technology Management, Holon Institute of Technology, Holon, Israel

[2]Faculty of Digital Technologies in Medicine, Holon Institute of Technology, Holon, Israel

[3]Faculty of Sciences, Holon Institute of Technology, Holon, Israel

[4]Adelson School of Medicine, Ariel University, Ariel, Israel

Corresponding Author:
Arriel Benis, PhD
Faculty of Industrial Engineering and Technology Management
Holon Institute of Technology
Golomb St. 52
Holon, 5810201
Israel
Phone: 972 3 5026892
Email: arrielb@hit.ac.il

Abstract

Background: Vaccines are one of the most important achievements of modern medicine. However, their acceptance is only partial, with vaccine hesitancy and refusal representing a major health threat. Influenza vaccines have low compliance since repeated, annual vaccination is required. Influenza vaccines stimulate discussions both in the real world and online. Social media is currently a significant source of health and medical information. Elucidating the association between social media engagement and influenza vaccination is important and may be applicable to other vaccines, including ones against COVID-19.

Objective: The goal of this study is to characterize profiles of social media engagement regarding the influenza vaccine and their association with knowledge and compliance in order to support improvement of future web-associated vaccination campaigns.

Methods: A weblink to an online survey in Hebrew was disseminated over social media and messaging platforms. The survey answers were collected during April 2020. Anonymous and volunteer participants aged 21 years and over answered 30 questions related to sociodemographics; social media usage; influenza- and vaccine-related knowledge and behavior; health-related information searching, its reliability, and its influence; and COVID-19-related information searching. A univariate descriptive data analysis was performed, followed by multivariate analysis via building a decision tree to define the most important attributes associated with vaccination compliance.

Results: A total of 213 subjects responded to the survey, of whom 207 were included in the analysis; the majority of the respondents were female, were aged 21 to 40 years, had 1 to 2 children, lived in central Israel, were secular Israeli natives, had higher education, and had a salary close to the national average. Most respondents (128/207, 61.8%) were not vaccinated against influenza in 2019 and used social media. Participants that used social media were younger, secular, and living in high-density agglomerations and had lower influenza vaccination rates. The perceived influence and reliability of the information on social media about COVID-19 were generally similar to those perceptions about influenza.

Conclusions: Using social media is negatively linked to compliance with seasonal influenza vaccination in this study. A high proportion of noncompliant individuals can lead to increased consumption of health care services and can, therefore, overload these health services. This is particularly crucial with a concomitant outbreak, such as COVID-19. Health care professionals should use improved and targeted health communication campaigns with the aid of experts in social media. Targeted communication, based on sociodemographic factors and personalized social media usage, might increase influenza vaccination rates and compliance with other vaccines as well.

(J Med Internet Res 2021;23(3):e25977) doi:10.2196/25977

KEYWORDS

influenza; vaccines; vaccination; social media; online social networking; health literacy; eHealth; information dissemination; access to information; COVID-19

Introduction

Background

Influenza is an acute, viral infectious disease characterized by high fever, cough, runny nose, muscle pain, joint pain, and severe exhaustion [1]. It is associated with multiple complications, including hepatitis, encephalitis, muscle tissue destruction, renal impairment, and secondary bacterial infection (ie, pneumonia, sinusitis, and, in children, middle ear infection) [2]. Although vaccines are one of the most important achievements of modern medicine, their acceptance among the population is only partial; this lack of compliance has been identified by the World Health Organization (WHO) as one of the major threats to public health. A wide variety of communication channels are now being used to improve worldwide vaccine responsiveness. There is a need to improve the efficacy of online communication, which is currently a significant source of health and medical information [3].

According to the WHO, every year around 1.5 billion people suffer from seasonal influenza, of whom 3 to 5 million have a serious illness, sometimes requiring hospitalization, and 650,000 die [1]. However, the percentage of the population who comply with seasonal vaccination against it is relatively low [4,5]. For example, in Israel, around 25% of the population are vaccinated against influenza each winter [6]. The second type of influenza relates to major changes in viral antigens occurring every few decades and, thus, leads to new influenza strains that are unfamiliar to the human immune system, causing a worldwide epidemic, called a pandemic, with tens of millions of sick patients and millions of deaths [1]. This occurred with the 1918 influenza pandemic, in which about one-third of the world's population was ill and 50 to 100 million people died [7-9].

Over the last decades, the internet has supported the monitoring, prediction, and surveillance (ie, infoveillance [10-12]) of epidemics and the behavior of the population, specifically in the context of influenza [13-15]. Furthermore, valuable information is discoverable, like early warnings of disease outbreaks, dissemination tracking, and resilience [16]. Social media and social networking services (SNSs) are powerful internet-based communication tools [17,18]. Each SNS has a variety of functionalities and goals. Facebook is a grand public-focused polyvalent platform. LinkedIn focuses on professional networking. Instagram, Flickr, and Pinterest deal with video and picture sharing. Twitter and Tumblr allow for blogging and microblogging. Reddit provides news aggregation, web content rating, and blogging services. Telegram is primarily used for instant and broadcast messaging to an unlimited number of subscribers over dedicated channels. Social media and SNSs, in particular, are also well known for disseminating evidence-based health care information and recommendations [14]. However, efficient and effective health-related information must be monitored and controlled for both *quality and reliability* [19], and *confidentiality, privacy, and ethics* of contacts between health care information customers and providers [20]. The social media impact must be understood in the field of health communication [21,22]. Accordingly, the development of relevant policies is needed [23,24] for reducing the risk and impacts of the *misinformation epidemic* or the *infodemic* spread on media, and for building the appropriate capacities to support eHealth and science literacy [25,26]. There is a crucial need for public health decision makers who are concerned with disease literacy to give the health information providers appropriate tools for efficiently disseminating the information, taking into account possible personal and environmental influences [12].

Vaccination against influenza is a significant and cost-effective protective mechanism for reducing the disease burden related to its morbidity and mortality [5]. Nevertheless, at the population level, its coverage is insufficient due to factors influencing vaccination decisions and hesitancy, such as risk-benefit misperception or accessibility to the health care system [27]. A major contribution to these factors is communication, involving both social and mass media, family, friends, and health care professionals [28]. More specifically, social media could affect the compliance of the population to vaccination guidelines [27]. For example, those who advocate against vaccines use social media to disseminate their messages on a large scale, increasing vaccine hesitancy or refusal in the population [29].

Concerning the new COVID-19 vaccines, evaluating the relationship between the population's perception of and compliance with the vaccine against influenza is important. Therefore, this will contribute to creating effective means of online communication to improve vaccine acceptance [30-33].

Aims and Objectives

Social media and SNSs have been used to improve vaccine response worldwide. However, they are also a forum for vaccine opponents and spreading of fake news. Understanding social media engagement, influence, and reliability is a critical point for improving the efficacy of advertising and publicity policies on social media. Our primary aim is to support the design and the implementation of future eHealth strategies and interventions on social media to increase the quality of targeted communication campaigns and the influenza vaccination rates [6,34-36]. Our main objective is to describe and characterize profiles regarding influenza vaccination and their association with social media engagement, influence, and reliability. We specifically focus on the Israeli population in this study. The findings of this research may then support vaccination campaigns against COVID-19.

This cross-sectional survey-based research is led by three hypotheses:

1. The use of social media influences the vaccination compliance of health care customers.
2. Influenza vaccination compliance is affected by social factors and by perceptions, reliability, and influence of information from social media.

3. The perceived influence and reliability of information from social media about SARS-CoV-2 or COVID-19 is similar to that regarding influenza.

Our goal is to identify sociodemographic and social media engagement attributes affecting influenza vaccination compliance. This research characterizes the differences between individuals vaccinated or not vaccinated against influenza during the 2019 season [37]. We attempt to understand whether there is a link between seasonal vaccination against influenza and social media engagement, influence, and perception of reliability during the COVID-19 pandemic.

This survey was granted ethical approval by the Ethics Committee of the Faculty of Technology Management of the Holon Institute of Technology (TM/2/2020/AB/002). The information provided by the participants during the survey are stored in a secured, encrypted manner, with restricted access provided by the institution of the principal researcher (AB).

Methods

Overview

We performed a cross-sectional survey of volunteers and anonymous Hebrew speakers over the internet about their online social network habits and their behavior concerning influenza vaccination. The survey was conducted over 14 days, between April 14 and 28, 2020, coinciding with the end of the seasonal influenza outbreak as well as the second month of the COVID-19 pandemic in Israel. Israel is a country in which a high percentage of the population uses the internet. With more than 80% of the population having at least one account on an SNS [38], it is among the highest in the world and is continuously increasing. Facebook is the leading social media platform used in Israel and the percentage of its users is continuously growing (eg, in April 2020: 62.87%; in August 2020: 86.47%) [39].

Upon consent, 213 participants were instructed to complete a 30-question survey asking about their usage and perception of health information available on social media, its reliability, and its influence on their compliance to vaccinate against influenza (see Multimedia Appendix 1) [40,41].

The survey was hosted on an Israeli website for the management of surveys in Hebrew—the IMKFORMS system—and its address was disseminated by publishing it on SNSs (ie, Facebook, Twitter, and LinkedIn) and instant messaging platforms (ie, WhatsApp and Telegram) and by sending its link via email to personal and professional contact lists.

The questionnaire included five subsets of questions, each with a specific focus:

1. Sociodemographics (9 questions), including gender, age range, relationship status, number of children, area of residence, country of birth, religious affiliation, education, and monthly income.
2. Social media usage (2 questions), including the self-estimated daily amount of social media use and the types of involvement on different social media platforms.
3. Influenza and vaccine-related knowledge and behavior (5 questions), including vaccine status in 2019, knowledge about the influenza vaccine, reasons for taking the seasonal influenza vaccine, and chronic disease in the family.
4. Health-related information searching and publishing, its reliability, and its influence (9 questions), including confidence in sources of information and searching for information, reliability and influence, and types and intensity of involvement on different social media platforms regarding health, vaccines, and influenza.
5. COVID-19- and vaccine-related information searching and publishing, its reliability, and its influence (5 questions), including sources and searching for information, reliability and influence, and types and intensity of involvement on different social media platforms regarding the COVID-19 pandemic.

The questions dealing with social media usage, reliability, and influence were in the form of matrix point rating multiple-choice questions with 4-point Likert scales.

Data Analysis

By using the exclusion criteria, we removed the full answer sets of responders who declared residence outside Israel or who did not answer at least one of the sociodemographic questions. We redefined some categories to facilitate the data analysis by working with groups comprising, as much as possible, the largest number of similar answers (eg, age groups, relationship status, number of children, area of residence, country of birth, education, and social media activity, reliability, and influence). Descriptive statistics such as frequencies and proportions were computed. Chi-square tests and Fisher exact tests were used to compare categorical variables. Cronbach α was used to measure the internal reliability of social media usage, reliability, and influence (Cronbach α=.949). The categorical variables were presented as numbers and percentages. Statistical significance was considered with a 2-sided *P* value of .05 or less. However, the borderline values have not been considered as not being significant in the evaluation [42]. To promote the effective focusing of communication to encourage vaccination against influenza, a set of factors were considered. Building a decision tree allowed, in this case, the definition of target profiles and the overcoming of Simpson's paradox, which may limit the quality of the decision support provided to decision makers. This phenomenon describes situations in which a trend appears in some groups of data but disappears when these groups are aggregated and vice versa [43]. Therefore, these learning classifiers allow nonlinear interactions between attributes and are easily interpretable. The decision tree for predicting vaccinated and unvaccinated profiles [44] was built by considering attributes with $P \leq .10$ that were used in further multivariate analysis.

The data analysis was performed with R, version 4.0.2 (The R Foundation). The *psych* package [45] was used for computing the internal consistency of reliability of the answers to the matrix multipoint scale questions. The *compareGroups* package [46] was used for statistical computations. The *rpart* package [47] was used for the decision tree processing [48].

The manuscript adheres to reporting standards, including the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [49-51].

Results

Study Population

Overview

The population of survey participants (see Table 1) includes 207 individuals after applying the exclusion criteria on 213 total responders. A substantial proportion of the participants were female (126/207, 60.9%), between the ages of 21 and 40 years (47/207, 71.0%), in a relationship (156/207, 75.4%), with 1 to 2 children (101/207, 48.8%), and living in central Israel (130/207, 62.8%). Additionally, the majority of survey participants were Israeli natives (116/207, 56.0%), had a secular affiliation (144/207, 69.6%), possessed higher education (182/207, 87.9%), and had a salary near the national average (81/207, 39.1%), which is around 12,000 New Israeli Shekels per month (US $3092.78) [52]. Regarding demography of social media usage and vaccination status, the majority of respondents were not vaccinated against influenza in 2019 (128/207, 61.8%) and used social media (119/207, 57.5%).

Table 1. Sociodemographic characteristics of the survey population.

Characteristic	All participants (N=207), n (%)	Social media user			Vaccinated in 2019		
		Yes (n=119), n (%)	No (n=88), n (%)	*P* value	Yes (n=79), n (%)	No (n=128), n (%)	*P* value
Vaccinated in 2019				<.001			N/A[a]
Yes	79 (38.2)	31 (26.1)	48 (55)	—[b]	N/A	N/A	N/A
No	128 (61.8)	88 (73.9)	40 (45)	—	N/A	N/A	N/A
Social media user				N/A			<.001
Yes	119 (57.5)	N/A	N/A	N/A	31 (39)	88 (68.8)	—
No	88 (42.5)	N/A	N/A	N/A	48 (61)	40 (31.2)	—
Gender				.07			.11
Male	81 (39.1)	53 (44.5)	28 (32)	—	25 (32)	56 (43.8)	—
Female	126 (60.9)	66 (55.5)	60 (68)	—	54 (68)	72 (56.2)	—
Age category (years)				.002			.11
21-30	48 (23.2)	34 (28.6)	14 (16)	—	12 (15)	36 (28.1)	—
31-40	99 (47.8)	63 (52.9)	36 (41)	—	41 (52)	58 (45.3)	—
41-50	38 (18.4)	14 (11.8)	24 (27)	—	15 (19)	23 (18.0)	—
51-60	15 (7.2)	7 (5.9)	8 (9)	—	6 (8)	9 (7.0)	—
≥61	7 (3.4)	1 (0.8)	6 (7)	—	5 (6)	2 (1.6)	—
Relationship status				.04			.008
Not in a relationship	51 (24.6)	36 (30.3)	15 (17)	—	11 (14)	40 (31.2)	—
In a relationship	156 (75.4)	83 (69.7)	73 (83)	—	68 (86)	88 (68.8)	—
No. of children				.009			.03
0	58 (28.0)	43 (36.1)	15 (17)	—	14 (18)	44 (34.4)	—
1-2	101 (48.8)	53 (44.5)	48 (55)	—	45 (57)	56 (43.8)	—
3-6	48 (23.2)	23 (19.3)	25 (28)	—	20 (25)	28 (21.9)	—
Residence				.02			.53
Center	130 (62.8)	66 (55.5)	64 (73)	—	47 (59)	83 (64.8)	—
Periphery	77 (37.2)	53 (44.5)	24 (27)	—	32 (41)	45 (35.2)	—
Country of birth				.002			.82
Israel	116 (56.0)	55 (46.2)	61 (69)	—	43 (54)	73 (57.0)	—
Aboard	91 (44.0)	64 (53.8)	27 (31)	—	36 (46)	55 (43.0)	—
Religious affiliation				.21			.03
Secular	144 (69.6)	89 (74.8)	55 (63)	—	50 (63)	94 (73.4)	—
Traditional	40 (19.3)	19 (16.0)	21 (24)	—	15 (19)	25 (19.5)	—
Religious	17 (8.2)	7 (5.9)	10 (11)	—	12 (15)	5 (3.9)	—
Other	6 (2.9)	4 (3.4)	2 (2)	—	2 (3)	4 (3.1)	—
Education (years)				.71			.99
≤12	25 (12.1)	13 (10.9)	12 (14)	—	9 (11)	16 (12.5)	—
>12	182 (87.9)	106 (89.1)	76 (86)	—	70 (89)	112 (87.5)	—
Monthly gross income (NIS[c])				.28			.75
5001-10,000	33 (15.9)	24 (20.2)	9 (10)	—	10 (13)	23 (18.0)	—
10,001-20,000	81 (39.1)	42 (35.3)	39 (44)	—	35 (44)	46 (35.9)	—

Characteristic	All participants (N=207), n (%)	Social media user			Vaccinated in 2019		
		Yes (n=119), n (%)	No (n=88), n (%)	*P* value	Yes (n=79), n (%)	No (n=128), n (%)	*P* value
20,001-25,000	28 (13.5)	16 (13.4)	12 (14)	—	9 (11)	19 (14.8)	—
25,001-30,000	20 (9.7)	14 (11.8)	6 (7)	—	7 (9)	13 (10.2)	—
≥30,001	25 (12.1)	13 (10.9)	12 (14)	—	9 (11)	16 (12.5)	—
Did not disclose	20 (9.7)	10 (8.4)	10 (11)	—	9 (11)	11 (8.6)	—
Do you or a family member[d] have a chronic disease?				.90			.45
Yes	94 (45.4)	55 (46.2)	39 (44)	—	39 (49)	55 (43.0)	—
No	113 (54.6)	64 (53.8)	49 (56)	—	40 (51)	73 (57.0)	—

[a]N/A: not applicable; irrelevant test.

[b]*P* values were calculated for categories and not for individual subcategories.

[c]NIS: New Israeli Shekel; the currency exchange rate at the time of publication was US $1 for NIS 3.88.

[d]Up to a second degree.

Social Media Users

There were significantly fewer social media users vaccinated against influenza in 2019 (31/109, 26.1%; *P*<.001) compared to nonusers (48/88, 55%). In both groups, the responders were predominantly female (users: 66/119, 55.5% vs nonusers: 60/88, 68%; *P*=.09) and in a relationship (83/119, 69.7% vs 73/88, 83%; *P*=.04). However, the social media users were globally younger, 21 to 40 years old (97/119, 81.5% vs 50/88, 67%), whereas the nonusers were older, aged between 31 and 60 years (68/88, 77%). Similarly, the former group often had no children (36/118, 80.6%), whereas the nonusers had children in most cases (73/88, 83%). Still, most respondents lived in central Israel (66/119, 55.5% and 64/88, 73%), but residents of the periphery used significantly more social media than those living in the center (53/119, 44.5% vs 24/88, 27%; *P*=.02). Additionally, in the survey population, Israeli natives used social media significantly less than immigrants (55/119, 46.2% vs 64/119, 53.8%; *P*=.002).

Furthermore, the proportion of people of traditional or religious affiliation who did not use social media was higher than their proportion in the group of users (31/88, 35% vs 26/119, 21.8%; *P*=.21). The level of education, declared monthly gross income, and experience of chronic disease by themselves or by a family member did not appear to be a determinant of overall social media use (*P*=.28, *P*=.71, and *P*=.90, respectively).

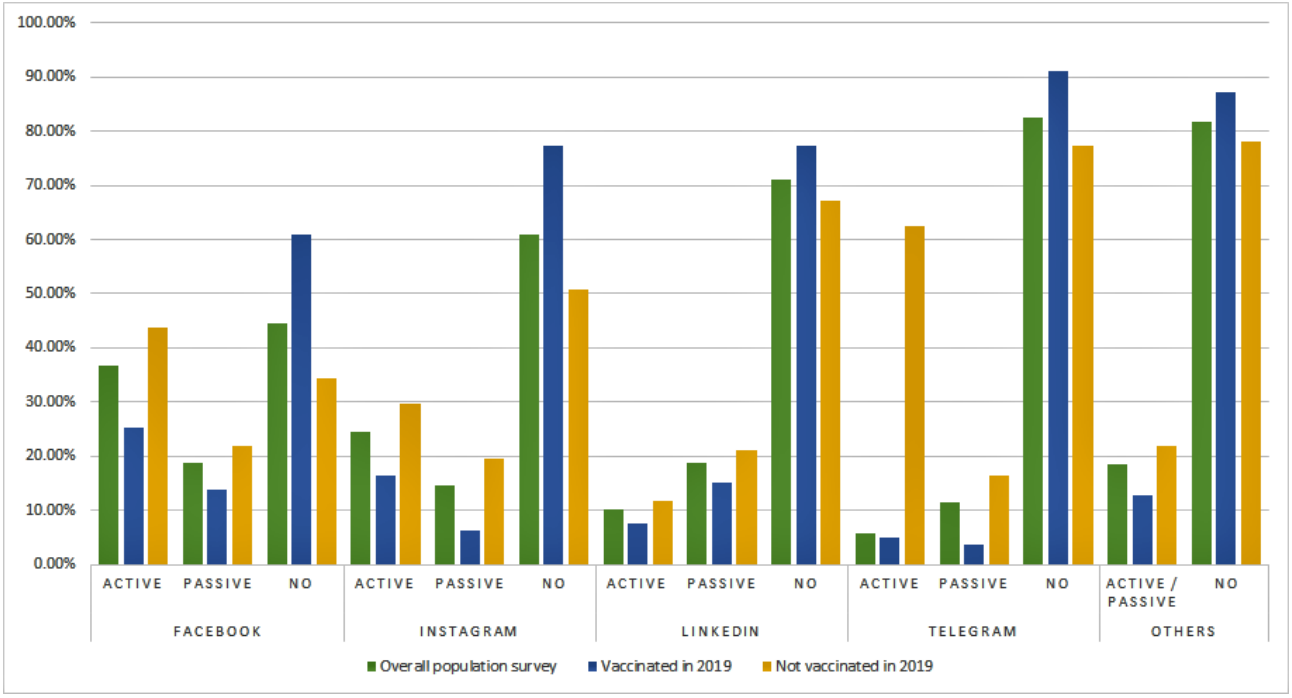
Vaccinated Versus Unvaccinated in 2019

The percentage of respondents vaccinated against influenza in 2019 was significantly (*P*<.001) higher in the group of social media nonusers than in the users' group (48/79, 61% vs 31/79, 39%); inversely, the percentage of unvaccinated respondents was higher in the group of social media users (88/128, 68.8% vs 40/128, 31.2%). Furthermore, the proportion of vaccinated people was higher when the responders were in a relationship (68/79, 86% vs 88/128, 68.8%; *P*=.008) and had children (65/79, 83% vs 84/128, 65.7%; *P*=.03). The degree of religious affiliation provided critical insight into this study. The secular participants in the survey represented a higher proportion of unvaccinated people (unvaccinated: 94/128, 73.4% vs vaccinated: 50/79, 63%), while those with religious affiliations were more compliant (unvaccinated: 5/128, 3.9% vs vaccinated: 12/79, 15%). Similar to the use of social media, influenza vaccination was not associated with chronic disease, personally or among family members (*P*=.45).

Social Media Usage and Vaccination Status

The participants in the survey were asked to specify which social media platforms they used actively (ie, publishing or reacting to posts), passively (ie, reading posts), or not at all (see Figure 1 and Multimedia Appendix 2). The participants declared using Facebook (115/207, 55.0%), Instagram (81/207, 39.1%), LinkedIn (60/207, 28.9%), Telegram (36/207, 17.4%), and others (ie, Twitter, Tumblr, Reddit, Flickr, and other social media platforms) as a whole (38/207, 18.4%). More accurately, Facebook, Instagram, and Telegram users were significantly (*P*=.001) less vaccinated in 2019. The use of LinkedIn or other unspecified social media platforms was not significantly associated with vaccination status (*P*=.30 and *P*=.14, respectively).

Figure 1. Social media usage and vaccination status against influenza in 2019.

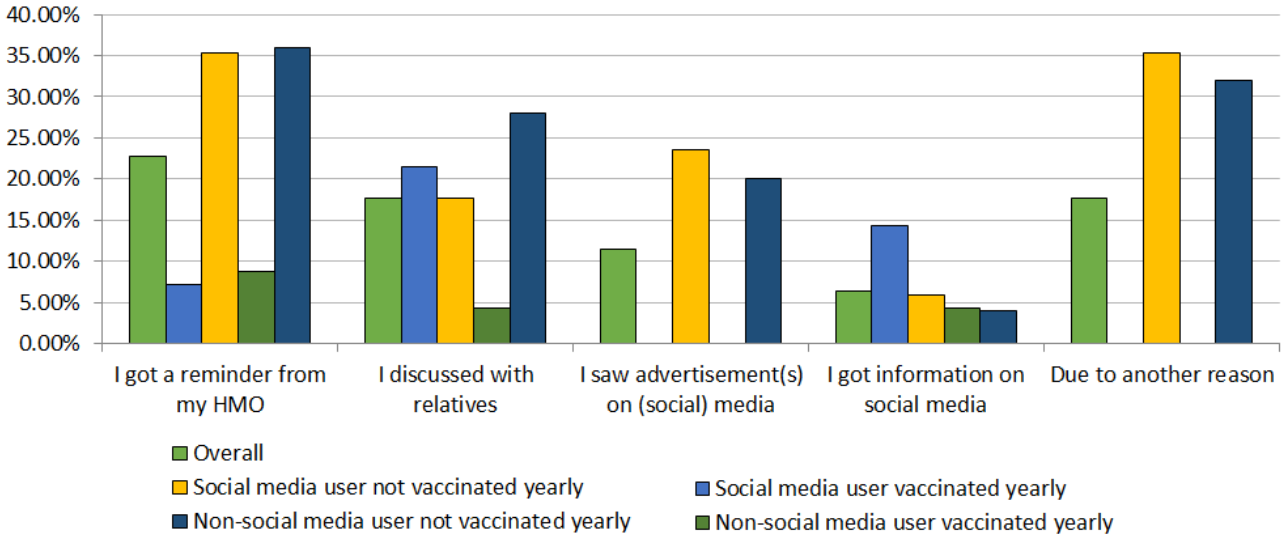


Moreover, no significant difference was observed between the groups regarding the belief that the vaccine against influenza caused the disease (between 62.0% and 68.0%; *P*>.47), nonetheless a majority of participants knew that the influenza vaccine is an attenuated or inactivated virus (150/207, 72.5%).

Reasons for Receiving the Influenza Vaccine in 2019

The reasons for receiving the annual influenza vaccine varied from one individual to another (see Figure 2 and Multimedia Appendix 3). The participants vaccinated in 2019 (79/207, 38.2%) were divided into *social media users* (31/79, 39%) and *nonusers of social media* (48/79, 61%). The proportion of yearly vaccinated individuals was similar in both groups (14/31, 45% vs 23/48, 48%). Receiving a reminder from a health maintenance organization (HMO) seemed to significantly influence the compliance with vaccination for the group of individuals who were not vaccinated annually (6/7, 86%; *P*=.09 vs 9/11, 82%; *P*=.06). Among those who have discussed the issue with relatives, seen advertisements in the media, or gotten information via social media, some differences were noticeable. The social media users seeing advertisements may have been influenced by these communication tools and took the vaccine in 2019 in contrast to previous years (4/9, 44%). A relatively high proportion (14/79, 18%) of individuals were vaccinated in 2019 but had not been vaccinated annually. This may be associated with the COVID-19 pandemic and the worry that it induced [53]. Moreover, a majority of the responders who were not vaccinated yearly received the vaccine after a discussion with relatives when they were not social media users (7/8, 88% vs 3/6, 50%; odds ratio 7.0, 95% CI 0.50-97.75; *P*=.15).

Figure 2. Reasons for receiving the influenza vaccine in 2019. The plot shows all participants and stratification by social media use and vaccination against influenza in 2019. HMO: health maintenance organization.



Searching for and Publishing Information Related to Health, Specifically to Influenza Vaccines

Social media is the main source of information and news consumption [54,55]. Anyone can post content and thus publish "information" (see Multimedia Appendix 4). No significant differences were observed in health information–related searches, influenza vaccine–related posting, and COVID-19-related posting (P>.30). However, declared behavior was significantly associated with vaccination against influenza and with the search for influenza vaccine– and COVID-19-related information (22/79, 28% vs 11/128, 8.6%; P<.001, and 26/79, 33% vs 23/128, 18.0%; P=.02, respectively), as well as publishing of health-related information (46/79, 58% vs 56/128, 43.8%; P=.06). These results show that vaccinated individuals were more active and involved in their health management by searching for and sharing relevant information.

Reliability, Influence, and Confidence Perceptions of Influenza-, Vaccine-, and COVID-19-Related Information on Social Media Platforms

The survey participants were asked to report their perceptions of reliability, influence, and confidence of influenza-, vaccine-, and COVID-19-related information available on the social media platforms (see Multimedia Appendices 5 and 6). The proportion of participants with *no opinion* was relatively high (at least 118/207, 57.0%) and higher than that in the *social media nonusers* group (88/207, 42.5%). The *no opinion* answer was not considered as a full lack of positioning but rather as a lack of use, knowledge, or understanding of a platform. It seems that the social media users were not aware of the impacts of these platforms on their behaviors. Facebook received the highest score of reliability (71/207, 34.3%; P=.06 and 74/207, 35.7%; P=.003) and the score was higher in the *nonvaccinated* group (50/128, 39.1%). The influence of the information about the influenza vaccine was not considered as being substantial (53/207, 25.6%). For COVID-19, the results were different: two-thirds of the participants, not vaccinated and having an opinion, were influenced by the information appearing on social media.

The users' trust in the source of information is crucial. The majority of participants had confidence in *governmental and health organizations* (overall: 155/207, 74.9%; vaccinated: 65/79, 82%; nonvaccinated: 38/128, 29.7%). Interestingly, the participants generally had less confidence in health care professionals as a source of influenza vaccine–related information (overall: 84/207, 40.6%; vaccinated: 39/79, 49%; nonvaccinated: 45/128, 35.2%; P=.06). These two sources represent those with the highest levels of trust and, as such, the highest levels of influence. One-third of the participants were confident in scientific publications (70/207, 33.8%). Moreover, the information provided by relatives had some credibility (17/207, 8.2%). The pharmaceutical industry and the vaccine opponents' information were strongly rejected (>97.6%).

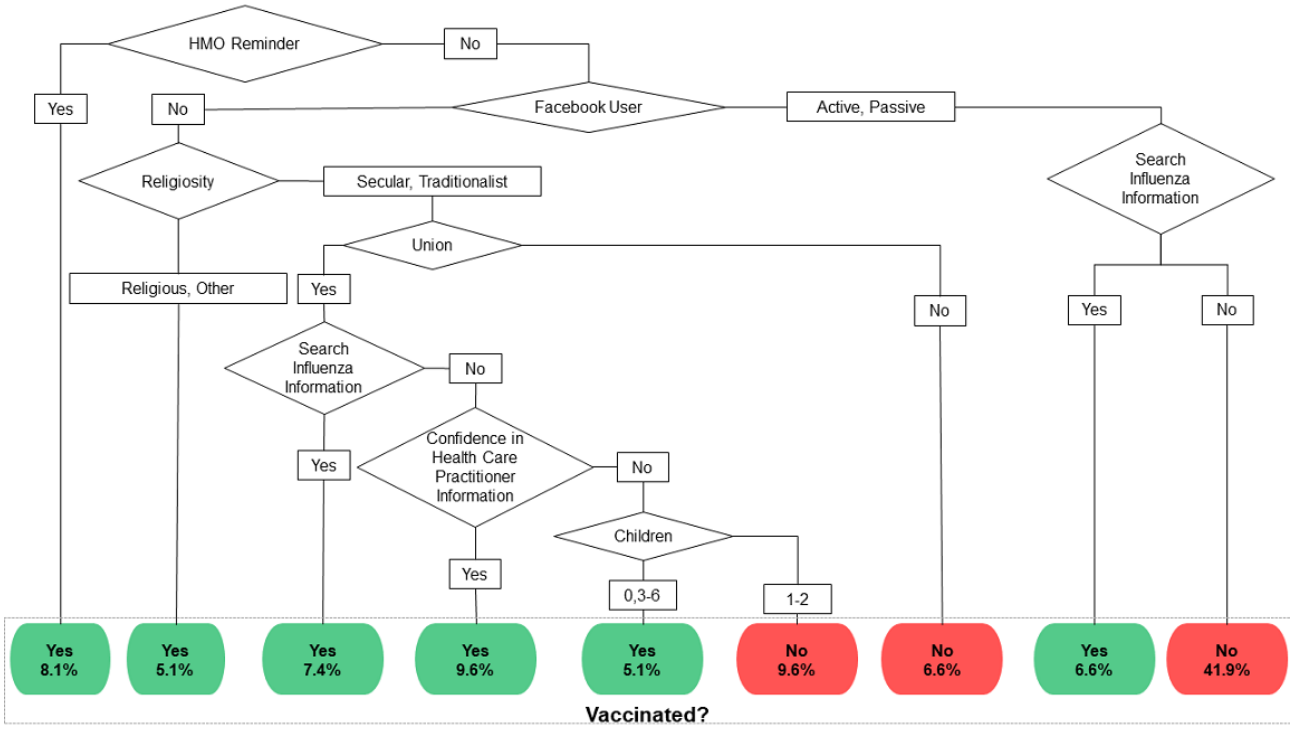
Multivariate Analysis

The multivariate analysis consisted of building decision trees enabling the classification of individuals to be vaccinated or unvaccinated against influenza and so defining target profiles for increasing vaccination engagement. As presented in the univariate analysis (Multimedia Appendices 2-7), multiple decision trees were built with the subsets of attributes (see Multimedia Appendix 8) without considering the P values, with and without using the COVID-19-related set. The decision trees were built using a training data set of 66.2% (137/207) randomly selected records and their prediction capabilities were tested on the rest (70/207, 33.8%).

One decision tree was built with all attributes with P≤.10 and with the exclusion of the COVID-19-related attributes (56/70, 80% of the test samples were rightly classified as vaccinated or not). The root of this decision tree was *due to another reason* as the reason given for being vaccinated in the 2019 season. In the overall survey, this represented 18% (14/79) of the vaccinated individuals or 6.8% of the responders (14/207). Moreover, the same group reported that they were not vaccinated annually (see Multimedia Appendix 3). To build a decision tree that supports efficient decision making by the domain expert, this attribute did not allow the targeting of a specific subpopulation. Hence, the increase in the number of people vaccinated may be due to fear of the pandemic [53] and the popular misunderstanding of the differences between seasonal influenza and SARS-CoV-2 (ie, the COVID-19 pandemic) [33].

Consequently, two decision trees were built without the *due to another reason* attribute: the first was built on all attributes with P≤.10, and the second further excluded the COVID-19-related attributes as inputs. The same decision tree was generated (see Figure 3), with an overall performance of 76% (53/70 of the test sample were rightly classified as vaccinated or not). The difference in the classification performances of the trees with the declaration *due to another reason* or without it was 4% (3/70). This means that the majority (11/14, 79%) of the individuals vaccinated for this *reason* were rightly classified even without taking it into account.

Figure 3. Decision tree predicting vaccinated individuals based on all survey attributes with $P \leq .10$ and without the *due to another reason* attribute. HMO: health maintenance organization.



Not focusing particularly on COVID-19 did not have an impact on classification. This supports our main objective, which was to characterize profiles of vaccination engagement in a simple and generalizable way.

The root of the proposed decision tree (see Figure 3) was a reason for obtaining the seasonal vaccine and it was a nonsocial attribute: *I got a reminder from my HMO* (11/137, 8.1% of the training set; P=.08). This reminder was sent by HMOs via SMS and the majority of those who received it were compliant and vaccinated (see Figure 3 and Multimedia Appendix 4).

The next node involved discrimination between Facebook users and nonusers (node *Facebook User*). The group of individuals engaged with this social media platform were less vaccinated than its nonusers. However, they actively searched for information related to influenza vaccines (ie, *Search Influenza Information*); thus, their health literacy seemed to positively influence vaccination compliance (9/70, 7%; P=.07).

On the other hand, the Facebook nonusers were then split by their religious practice (ie, *Religiosity*). The individuals who were not vaccinated because they did not receive an HMO reminder and were also not users of Facebook were, for the majority, vaccinated if they were *religious* or had another level of practice, meaning neither *secular* nor *traditional* (7/137, 5.1%; P=.05). In the last node, those not in a relationship were more likely not to be vaccinated (9/137, 6.6%; P=.07).

The majority of those who identified as secular or traditional, were in a relationship, and searched for influenza vaccine information were also vaccinated (10/137, 7.4%; P=.07). Those who did not search for information but were confident in the information provided by health care practitioners also showed good compliance (13/137, 9.6%; P=.10). Surprisingly, those who were not confident, and who had no children or more than 3 children, were often vaccinated (7/137, 5.1%; P=.05), while parents of 1 or 2 children were less often vaccinated (13/137, 9.6%; P=.10).

Discussion

Principal Findings

This study aimed to quantify the contribution of social media and its perceived reliability to modern health care customer behavior, thus highlighting the centrality of these media platforms to influence treatment and, specifically, compliance with influenza vaccination [56,57]. The most important outcome of this study is building a decision tree that is based on our findings and supports a multivariate and integrative viewpoint. It is actually reasonable that various interacting links exist between the seasonal vaccination against influenza and social media engagement, influence, and reliability in Israel during the COVID-19 pandemic [39]. Indeed, being or not being a user of social media or searching for information about the influenza vaccine present strong associations with vaccination status. Nevertheless, the vaccination reminders sent by the HMO are one of the most crucial factors in vaccination compliance.

The results show that the use of social media influences vaccination compliance in the Israeli population. This is correlated with sociodemographic factors and with perceptions of influence and reliability of information from social media. Indeed, social media users were less frequently vaccinated than nonusers. More accurately, this group was composed of younger people with secular affiliations living in the center of the country (ie, high-density agglomeration) and were less vaccinated.

According to the data collected during the survey, social media is largely used in Israel by young (ie, <40 years old), urban (ie, specifically, living in densely populated central Israel), and

more highly educated people. This research highlights that social media users and the majority of the nonvaccinated population are the younger population who also acknowledge their confidence in the reliability and influence of social media, concurrent with their low confidence in the information disseminated by governmental and health organizations and health professionals.

The perceived influence and reliability of the information on social media about COVID-19 are similar, in general, to those perceptions of the information about influenza. Accordingly, the risk of nonadherence to recommendations by governmental and health organizations to reduce the spread of these pandemic viruses (ie, hygiene measures, distancing, and vaccination) may be similar. Notably, while both parameters are relatively significant, the nonvaccinated subpopulation searches for information regarding influenza vaccination substantially less than information regarding SARS-CoV-2 or COVID-19. Social media platforms were used during the survey period by local, national, and international agencies (ie, governmental and health care organizations) to provide up-to-date facts, guidance, and directives to the public [58]. With this in mind, and considering the 2019-2020 influenza season time frame [37], 6.8% (14/207) of the respondents sought out the influenza vaccine, deviating from their habits.

Strengths and Limitations

A major limitation of this study is that it was based on an online survey written in Hebrew and disseminated over the internet via social media, primarily to Israeli residents. This method limited the type of individuals who responded to the survey and the generalization of the results. Moreover, even though about half of the survey responders were not Israeli natives, the Hebrew language of the survey rendered it inaccessible to nonfluent Hebrew speakers (eg, new immigrants, residents who have not learned the language, and non-Hebrew speakers of the Arab sector). Furthermore, as the survey was disseminated over the internet, particularly by publication on social networks, it did not include representation of the ultra-Orthodox sector, which represents around 12% of the Israeli population. It is, therefore, suggested that similar studies be performed in other locations and using diverse languages. Still, Israel is recognized as one of the countries with high, and continuously increasing, penetration of internet and social media [38,39].

Regarding the strengths of the study, the seasonal influenza outbreaks and pandemics with the need for repeated vaccinations are an excellent model for understanding the behavior of the population, its risk perception [59], its fear [60], and the consequences of information fatigue [61]. Social media supports the dissemination of a constant flow of information from numerous sources. In the context of influenza and COVID-19, these sources are institutional and professional news channels, in parallel with mass populations who can easily share their opinions and information over any social media platform. Crisis-related communication uses social media to “communicate, self-organize, manage, and mitigate risks” and “make sense of the event” [62] in a rapid manner and on influential channels [63].

Indeed, social media and the internet, in general, are defined as common sources of information on measles and its vaccine. Similar to the findings of this study, use of social media is also associated with erroneous knowledge and noncompliance with vaccination recommendations against measles [64].

Health communication countermeasures must be developed to increase the efficiency of campaigns for vaccination [64]. They must be dynamically adapted over all the communication channels in order to increase engagement for influenza and other vaccines, thereby reducing the potential additional overload on health care organizations [65,66].

Conclusions

Social media is currently a leading and user-centered source of health information. The information available about influenza and the vaccines against it appears reliable and influences its readers, and this is significantly correlated with seasonal influenza vaccination compliance. Accordingly, it is crucial to improve the targeting of health communication campaigns in social media in order to increase compliance.

A high proportion of noncompliant and, therefore, nonvaccinated individuals can lead to increased consumption of health care services and an overload of the system. Therefore, in the COVID-19 era, and similarly to other epidemics and pandemics, health care services are overwhelmed by an excess of sick patients. Consequently, efficient communication actions with the individual, familial, and societal spheres; education on the benefits of vaccination; and education on the risks associated with infection are essential.

A future intervention must be efficiently implemented by developing new health communication processes, considering vaccinated and nonvaccinated individual profiles defined in a periodically updated way by building, for example, specific decision trees as proposed in this research (see Figure 3). This must be based on improving the health communication flow on social media, in near-real time, by monitoring and adapting targeted campaigns for facing rapid information changes in the internet sphere (eg, breaking news and fake news) [67].

Furthermore, most of the participants in this research were *young*. Nowadays, social media and social networks are used across the population regardless of, for example, age, gender, relationship status, education, religious affiliation, and country of residence. Consequently, social media–based health communication aiming to increase treatment compliance and, more accurately herein, influenza vaccination must take into account sociodemographic variables at a local level for efficiently and effectively targeting individuals to be motivated to vaccinate. This means delivering information to social media users in language that is easy to understand, in the languages spoken by various native and immigrant communities in the country, to ensure that the information reaches all residents. Additionally, the use of reminders, similar to those on SMS, should be generalized to social media by utilizing tailored advertisements [68,69].

Acknowledgments

The research of AB, EL, and SA is supported by a grant from Ariel University and Holon Institute of Technology, Israel: *Applications of Artificial Intelligence for Enhancing the Efficiency of Vaccination Programs* (No. RA19000000649).

Authors' Contributions

All authors attest that they meet the International Committee of Medical Journal Editors criteria for authorship, have reviewed the version of the manuscript to be submitted, and agreed with its content and submission. AB was responsible for the conception, design, and conduct of the study; data analysis; preparation and submission of documentation to the Ethics Committee; data interpretation; writing of the first draft of the manuscript; and critical review and revision of the manuscript for important intellectual content. AN and SR were responsible for the conception and design of the study, data collection, data analysis, data interpretation; writing of the first draft of the manuscript; and critical review and revision of the manuscript for important intellectual content. AN and SR are students of the Master of Science of Technology Management, Holon Institute of Technology, Israel (under the supervision of AB). EL was responsible for data analysis, data interpretation, and critical review and revision of the manuscript for important intellectual content. SA was responsible for the conception and design of the study, data analysis, data interpretation, and critical review and revision of the manuscript for important intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire on the behavior of social media users regarding health and vaccine issues.
[PDF File (Adobe PDF File), 81 KB - jmir_v23i3e25977_app1.pdf]

Multimedia Appendix 2
Social media general usage and vaccination status against influenza in 2019.
[PDF File (Adobe PDF File), 62 KB - jmir_v23i3e25977_app2.pdf]

Multimedia Appendix 3
Reasons for obtaining the influenza vaccine in 2019. The plot shows all participants and stratification by social media use and vaccination against influenza in 2019.
[PDF File (Adobe PDF File), 59 KB - jmir_v23i3e25977_app3.pdf]

Multimedia Appendix 4
Searching for and publishing information related to health and specifically to influenza vaccines and COVID-19 in the 12 months before participating in the survey.
[PDF File (Adobe PDF File), 59 KB - jmir_v23i3e25977_app4.pdf]

Multimedia Appendix 5
Perception of reliability and influence of the information related to influenza and vaccine and COVID-19 available on the most used social media platforms.
[PDF File (Adobe PDF File), 73 KB - jmir_v23i3e25977_app5.pdf]

Multimedia Appendix 6
Perception of reliability and influence of the information related to influenza and vaccine and COVID-19 available on the most used social media platforms between social media users and social media nonusers.
[PDF File (Adobe PDF File), 64 KB - jmir_v23i3e25977_app6.pdf]

Multimedia Appendix 7
Confidence in sources of information about the vaccine against influenza.
[PDF File (Adobe PDF File), 51 KB - jmir_v23i3e25977_app7.pdf]

Multimedia Appendix 8
Prediction capabilities of decision trees based on the survey data.
[PDF File (Adobe PDF File), 67 KB - jmir_v23i3e25977_app8.pdf]

References

1. Influenza (Seasonal). World Health Organization. 2018 Nov 06. URL: https://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal) [accessed 2021-03-09]
2. Rothberg MB, Haessler SD. Complications of seasonal and pandemic influenza. Crit Care Med 2010 Apr;38(4 Suppl):e91-e97 [FREE Full text] [doi: 10.1097/CCM.0b013e3181c92eeb] [Medline: 19935413]
3. Rivas H, Wac K, editors. Digital Health: Scaling Healthcare to the World. Cham, Switzerland: Springer International Publishing; 2018.
4. Marshall GS. The Vaccine Handbook: A Practical Guide for Clinicians. 3rd edition. West Islip, NY: Professional Communications, Inc; 2010.
5. Fine P, Eames K, Heymann DL. "Herd immunity": A rough guide. Clin Infect Dis 2011 Apr 01;52(7):911-916 [FREE Full text] [doi: 10.1093/cid/cir007] [Medline: 21427399]
6. Feemster KA. Building vaccine acceptance through communication and advocacy. Hum Vaccin Immunother 2020 May 03;16(5):1004-1006. [doi: 10.1080/21645515.2020.1746603] [Medline: 32401681]
7. Taubenberger JK, Morens DM. 1918 influenza: The mother of all pandemics. Emerg Infect Dis 2006 Jan;12(1):15-22 [FREE Full text] [doi: 10.3201/eid1201.050979] [Medline: 16494711]
8. Potter CW. A history of influenza. J Appl Microbiol 2001 Oct;91(4):572-579 [FREE Full text] [doi: 10.1046/j.1365-2672.2001.01492.x] [Medline: 11576290]
9. Quinn T. Flu: A Social History of Influenza. London, UK: New Holland Publishers; 2008.
10. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):e20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
11. Eysenbach G. Infodemiology: The epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765 [FREE Full text] [doi: 10.1016/s0002-9343(02)01473-0] [Medline: 12517369]
12. Eysenbach G. Infodemiology and infoveillance: Tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158 [FREE Full text] [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
13. Corley C, Cook D, Mikler A, Singh K. Using web and social media for influenza surveillance. Adv Exp Med Biol 2010;680:559-564 [FREE Full text] [doi: 10.1007/978-1-4419-5913-3_61] [Medline: 20865540]
14. Alessa A, Faezipour M. A review of influenza detection and prediction through social networking sites. Theor Biol Med Model 2018 Feb 01;15(1):2 [FREE Full text] [doi: 10.1186/s12976-017-0074-5] [Medline: 29386017]
15. Yuan X, Crooks A. Examining online vaccination discussion and communities in Twitter. In: Proceedings of the 9th International Conference on Social Media and Society. New York, NY: Association for Computing Machinery; 2018 Presented at: 9th International Conference on Social Media and Society; July 18-20, 2018; Copenhagen, Denmark p. 197-206. [doi: 10.1145/3217804.3217912]
16. Mavragani A. Infodemiology and infoveillance: Scoping review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
17. McKee M, Cole K, Hurst L, Aldridge RW, Horton R. The other Twitter revolution: How social media are helping to monitor the NHS reforms. BMJ 2011 Feb 16;342:d948. [doi: 10.1136/bmj.d948] [Medline: 21325389]
18. van Dijck J. The Culture of Connectivity: A Critical History of Social Media. Oxford, UK: Oxford University Press; 2013.
19. Neumark Y, Flum L, Lopez-Quintero C, Shtarkshall R. Quality of online health information about oral contraceptives from Hebrew-language websites. Isr J Health Policy Res 2012 Sep 24;1(1):38 [FREE Full text] [doi: 10.1186/2045-4015-1-38] [Medline: 23006798]
20. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
21. Brodie M, Flournoy RE, Altman DE, Blendon RJ, Benson JM, Rosenbaum MD. Health information, the internet, and the digital divide. Health Aff (Millwood) 2000;19(6):255-265. [doi: 10.1377/hlthaff.19.6.255] [Medline: 11192412]
22. Neuhauser L, Kreps GL. Rethinking communication in the eHealth era. J Health Psychol 2003 Jan;8(1):7-23. [doi: 10.1177/1359105303008001426] [Medline: 22113897]
23. Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: A review and tutorial of applications in medicine and health care. J Med Internet Res 2014 Feb 11;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
24. Dixon G. Social media as a platform for science and health engagement: Challenges and opportunities. Isr J Health Policy Res 2016;5:57 [FREE Full text] [doi: 10.1186/s13584-016-0114-3] [Medline: 27895894]
25. Wu L, Morstatter F, Carley KM, Liu H. Misinformation in social media. SIGKDD Explor Newsl 2019 Nov 26;21(2):80-90. [doi: 10.1145/3373464.3373475]
26. Eysenbach G. How to fight an infodemic: The four pillars of infodemic management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
27. Ahmed N, Quinn SC, Hancock GR, Freimuth VS, Jamison A. Social media use and influenza vaccine uptake among White and African American adults. Vaccine 2018 Nov 26;36(49):7556-7561. [doi: 10.1016/j.vaccine.2018.10.049] [Medline: 30389192]

28. Schulz PJ, Wu Y, An S, Chen F. Social influences on influenza vaccination decision among senior citizens in Taiwan, South Korea, and Switzerland. Front Commun 2019 Dec 13;4:1-13 [FREE Full text] [doi: 10.3389/fcomm.2019.00074]
29. Maci SM. Discourse strategies of fake news in the anti-vax campaign. Lingue Culture Mediazioni 2019 Nov 12;6(1):15-43 [FREE Full text] [doi: 10.7358/lcm-2019-001-maci]
30. Barello S, Falcó-Pegueroles A, Rosa D, Tolotti A, Graffigna G, Bonetti L. The psychosocial impact of flu influenza pandemics on healthcare workers and lessons learnt for the COVID-19 emergency: A rapid review. Int J Public Health 2020 Sep;65(7):1205-1216 [FREE Full text] [doi: 10.1007/s00038-020-01463-7] [Medline: 32888048]
31. Jones DS. History in a crisis - Lessons for Covid-19. N Engl J Med 2020 Apr 30;382(18):1681-1683. [doi: 10.1056/NEJMp2004361] [Medline: 32163699]
32. Chong YY, Cheng HY, Chan HYL, Chien WT, Wong SYS. COVID-19 pandemic, infodemic and the role of eHealth literacy. Int J Nurs Stud 2020 Aug;108:103644 [FREE Full text] [doi: 10.1016/j.ijnurstu.2020.103644] [Medline: 32447127]
33. Solomon DA, Sherman AC, Kanjilal S. Influenza in the COVID-19 era. JAMA 2020 Oct 06;324(13):1342-1343. [doi: 10.1001/jama.2020.14661] [Medline: 32797145]
34. Massey PM, Leader A, Yom-Tov E, Budenz A, Fisher K, Klassen AC. Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter. J Med Internet Res 2016 Dec 05;18(12):e318 [FREE Full text] [doi: 10.2196/jmir.6670] [Medline: 27919863]
35. Henrich NJ. Increasing pandemic vaccination rates with effective communication. Hum Vaccin 2011 Jun;7(6):663-666. [doi: 10.4161/hv.7.6.15007] [Medline: 21445004]
36. Lama Y, Hu D, Jamison A, Quinn SC, Broniatowski DA. Characterizing trends in human papillomavirus vaccine discourse on Reddit (2007-2015): An observational study. JMIR Public Health Surveill 2019 Mar 27;5(1):e12480 [FREE Full text] [doi: 10.2196/12480] [Medline: 30916662]
37. 2019-2020 US flu season: Preliminary in-season burden estimates. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/flu/about/burden/preliminary-in-season-estimates.htm [accessed 2021-03-09]
38. Israel: Internet usage, broadband and telecommunications reports. Internet World Stats. 2020. URL: https://www.internetworldstats.com/me/il.htm [accessed 2021-03-09]
39. Social media stats: Israel. StatCounter. URL: https://gs.statcounter.com/social-media-stats/all/israel [accessed 2021-03-09]
40. Livni G, Chodik G, Yaari A, Tirosh N, Ashkenazi S. Attitudes, knowledge and factors related to acceptance of influenza vaccine by pediatric healthcare workers. J Pediatr Infect Dis 2015 Jul 28;03(02):111-117. [doi: 10.1055/s-0035-1556979]
41. Larson HJ, Jarrett C, Schulz WS, Chaudhuri M, Zhou Y, Dube E, SAGE Working Group on Vaccine Hesitancy. Measuring vaccine hesitancy: The development of a survey tool. Vaccine 2015 Aug 14;33(34):4165-4175 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.037] [Medline: 25896384]
42. Smith CJ, Fox ZV. The use and abuse of hypothesis tests: How to present P values. Phlebology 2010 Jun;25(3):107-112. [doi: 10.1258/phleb.2010.009094] [Medline: 20483859]
43. Shmueli G, Yahav I. The forest or the trees? Tackling Simpson's paradox with classification trees. Prod Oper Manag 2017 Dec 27;27(4):696-716. [doi: 10.1111/poms.12819]
44. Rokach L, Maimon O. Decision trees. In: Maimon O, Rokach L, editors. Data Mining and Knowledge Discovery Handbook. Boston, MA: Springer; 2005:165-192.
45. Using R and psych for personality and psychological research. Personality Project. 2020. URL: http://personality-project.org/r/psych/ [accessed 2021-03-09]
46. Subirana I, Sanz H, Vila J. Building bivariate tables: The compareGroups package for R. J Stat Softw 2014 May;57(12):1-16 [FREE Full text] [doi: 10.18637/jss.v057.i12]
47. rpart v4.1-15. RDocumentation. 2019. URL: https://www.rdocumentation.org/packages/rpart/versions/4.1-15 [accessed 2021-03-09]
48. Milborrow S. rpart.plot: Plot 'rpart' models: An enhanced version of 'plot.rpart'. The Comprehensive R Archive Network. 2020 Sep 17. URL: https://CRAN.R-project.org/package=rpart.plot [accessed 2021-03-09]
49. Eysenbach G. Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
50. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. PLoS Med 2007 Oct 16;4(10):e296 [FREE Full text] [doi: 10.1371/journal.pmed.0040296] [Medline: 17941714]
51. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. PLoS Med 2007 Oct 16;4(10):e297 [FREE Full text] [doi: 10.1371/journal.pmed.0040297] [Medline: 17941715]
52. Barkat A. Average salary in Israel falls to NIS 11,459. Globes. 2020 Aug 04. URL: https://en.globes.co.il/en/article-average-salary-in-israel-falls-to-nis-11459-1001338212 [accessed 2021-03-09]
53. Tvizer I. Health report: Worst flu epidemic in decade. Ynet News. 2020 Jan 06. URL: https://www.ynetnews.com/health_science/article/S1kM9axlI [accessed 2021-03-09]
54. Martin N. How social media has changed how we consume news. Forbes. 2018 Nov 30. URL: https://www.forbes.com/sites/nicolemartin1/2018/11/30/how-social-media-has-changed-how-we-consume-news/ [accessed 2021-03-09]

55. Brossard D. New media landscapes and the science information consumer. Proc Natl Acad Sci U S A 2013 Aug 20;110 Suppl 3:14096-14101 [FREE Full text] [doi: 10.1073/pnas.1212744110] [Medline: 23940316]
56. DeVito M, Birnholtz J, Hancock J. Platforms, people, and perception: Using affordances to understand self-presentation on social media. In: Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing. New York, NY: Association for Computing Machinery; 2017 Presented at: 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing; February 25-March 1, 2017; Portland, OR p. 740-754. [doi: 10.1145/2998181.2998192]
57. Benis A, Barak Barkan R, Sela T, Harel N. Communication behavior changes between patients with diabetes and healthcare providers over 9 years: Retrospective cohort study. J Med Internet Res 2020 Aug 11;22(8):e17186 [FREE Full text] [doi: 10.2196/17186] [Medline: 32648555]
58. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr;26(4):459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]
59. Barrios JM, Hochberg Y. Risk Perception Through the Lens of Politics in the Time of the COVID-19 Pandemic. Cambridge, MA: National Bureau of Economic Research; 2020 Apr. URL: https://www.nber.org/system/files/working_papers/w27008/w27008.pdf [accessed 2021-03-09]
60. Van Bavel JJ, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
61. Dhir A, Kaur P, Chen S, Pallesen S. Antecedents and consequences of social media fatigue. Int J Inf Manage 2019 Oct;48:193-202. [doi: 10.1016/j.ijinfomgt.2019.05.021]
62. Stieglitz S, Bunker D, Mirbabaie M, Ehnis C. Sense-making in social media during extreme events. J Conting Crisis Manage 2017 Aug 18;26(1):4-15 [FREE Full text] [doi: 10.1111/1468-5973.12193]
63. Ali KF, Whitebridge S, Jamal MH, Alsafy M, Atkin SL. Perceptions, knowledge, and behaviors related to COVID-19 among social media users: Cross-sectional study. J Med Internet Res 2020 Sep 08;22(9):e19913 [FREE Full text] [doi: 10.2196/19913] [Medline: 32841153]
64. Ashkenazi S, Livni G, Klein A, Kremer N, Havlin A, Berkowitz O. The relationship between parental source of information and knowledge about measles / measles vaccine and vaccine hesitancy. Vaccine 2020 Oct 27;38(46):7292-7298. [doi: 10.1016/j.vaccine.2020.09.044] [Medline: 32981777]
65. Shen K, Namazova-Baranova L, Yang Y, Wong GWK, Rosenwasser LJ, Rodewald LE, Global Pediatric Pulmonology Alliance (GPPA) Council, et al. Global Pediatric Pulmonology Alliance recommendation to strengthen prevention of pediatric seasonal influenza under COVID-19 pandemic. World J Pediatr 2020 Oct;16(5):433-437 [FREE Full text] [doi: 10.1007/s12519-020-00389-7] [Medline: 32920745]
66. Adlhoch C, Pebody R. What to expect for the influenza season 2020/21 with the ongoing COVID-19 pandemic in the World Health Organization European Region. Euro Surveill 2020 Oct;25(42):1-3 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.42.2001816] [Medline: 33094719]
67. Reukers DFM, Marbus SD, Smit H, Schneeberger P, Donker G, van der Hoek W, et al. Media reports as a source for monitoring impact of influenza on hospital care: Qualitative content analysis. JMIR Public Health Surveill 2020 Mar 04;6(1):e14627 [FREE Full text] [doi: 10.2196/14627] [Medline: 32130197]
68. Smailhodzic E, Hooijsma W, Boonstra A, Langley DJ. Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016 Aug 26;16:442 [FREE Full text] [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]
69. Shanahan T, Tran TP, Taylor EC. Getting to know you: Social media personalization as a means of enhancing brand loyalty and perceived quality. J Retailing and Consum Serv 2019 Mar;47:57-65. [doi: 10.1016/j.jretconser.2018.10.007]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
HMO: health maintenance organization
SNS: social networking service
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
WHO: World Health Organization

Edited by G Eysenbach; submitted 23.11.20; peer-reviewed by A Seidmann, DLG Rodrigues; comments to author 19.12.20; revised version received 28.12.20; accepted 01.03.21; published 16.03.21.

Please cite as:
Benis A, Khodos A, Ran S, Levner E, Ashkenazi S
Social Media Engagement and Influenza Vaccination During the COVID-19 Pandemic: Cross-sectional Survey Study
J Med Internet Res 2021;23(3):e25977
URL: https://www.jmir.org/2021/3/e25977
doi:10.2196/25977
PMID:33651709

Original Paper

Minimizing the Impact of the COVID-19 Epidemic on Oncology Clinical Trials: Retrospective Study of Beijing Cancer Hospital

Zhiying Fu[1*], MPH; Min Jiang[1*], M.Pharm; Kun Wang[1], MD; Jian Li[1], MD

Beijing Institute for Cancer Research, Beijing Cancer Hospital, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Jian Li, MD
Beijing Institute for Cancer Research
Beijing Cancer Hospital
No 52 Fucheng Road
Haidian District
Beijing
China
Phone: 86 1088196949
Email: LIJIAN8409@126.com

Abstract

Background: In view of repeated COVID-19 outbreaks in most countries, clinical trials will continue to be conducted under outbreak prevention and control measures for the next few years. It is very significant to explore an optimal clinical trial management model during the outbreak period to provide reference and insight for other clinical trial centers worldwide.

Objective: The aim of this study was to explore the management strategies used to minimize the impact of the COVID-19 epidemic on oncology clinical trials.

Methods: We implemented a remote management model to maintain clinical trials conducted at Beijing Cancer Hospital, which realized remote project approval, remote initiation, remote visits, remote administration and remote monitoring to get through two COVID-19 outbreaks in the capital city from February to April and June to July 2020. The effectiveness of measures was evaluated as differences in rates of protocol compliance, participants lost to follow-up, participant withdrawal, disease progression, participant mortality, and detection of monitoring problems.

Results: During the late of the first outbreak, modifications were made in trial processing, participant management and quality control, which allowed the hospital to ensure the smooth conduct of 572 trials, with a protocol compliance rate of 85.24% for 3718 participants across both outbreaks. No COVID-19 infections were recorded among participants or trial staff, and no major procedural errors occurred between February and July 2020. These measures led to significantly higher rates of protocol compliance and significantly lower rates of loss to follow-up or withdrawal after the second outbreak than after the first, without affecting rates of disease progression or mortality. The hospital provided trial sponsors with a remote monitoring system in a timely manner, and 3820 trial issues were identified.

Conclusions: When public health emergencies occur, an optimal clinical trial model combining on-site and remote management could guarantee the health care and treatment needs of clinical trial participants, in which remote management plays a key role.

(J Med Internet Res 2021;23(3):e26799) doi:10.2196/26799

KEYWORDS

COVID-19; clinical trials; management strategy; information technology

Introduction

COVID-19 was recognized as a global pandemic on March 11, 2020 [1,2]. Efforts to stem the spread of SARS-CoV-2 led to isolation measures and travel restrictions that have severely hindered the smooth conduct of clinical trials [3]. Participants may no longer be able to visit the hospital to receive treatment [4], resulting in protocol deviations. The regular monitoring and auditing of trials may also be affected [5], leading trials to be postponed or terminated; this in turn can delay marketing of drugs and even lead to the scrapping of development plans for new drugs. Several studies from several countries have reported obvious decreases in the numbers of participants newly enrolled

in trials, and many oncology clinical trials were suspended in April 2020 in several countries [6-8].

It seems likely that as COVID-19 outbreaks repeatedly occur in many countries, clinical trials will continue to be conducted under public health emergency conditions for some time to come. The conduct of oncology clinical trials will be particularly challenging, given that such trials usually require many follow-up visits and sophisticated monitoring of complex disease courses [9,10]. Therefore, insights into how to optimize oncology clinical trial management during public health emergencies are urgently needed.

This study examined the experiences of Beijing Cancer Hospital in its efforts to conduct anticancer drug trials through two COVID-19 outbreaks in 2020. During the first outbreak from February to April 2020, the hospital implemented a series of modifications in participant management and monitoring, including remote drug administration, remote visits, and remote trial monitoring. Relying on data processing and application platform (DPAP) technology, remote trial monitoring enables clinical research associates and inspectors to monitor trial data at any time and in any place, regardless of geographical and time constraints. At present, remote trial monitoring is rare in China and other countries. Here, we analyzed how the COVID-19 epidemic affected ongoing clinical trials and whether the hospital's modifications to trial management during the epidemic helped ensure the smooth conduct of these trials through the second outbreak from June to July 2020. Our experiences may be relevant to other countries as they seek to ensure the functioning of clinical trials during the COVID-19 epidemic and other public health emergencies.

Methods

This study was approved by the Ethics Committee of Beijing Cancer Hospital, which waived the requirement for informed consent for the participants because those individuals had already provided written informed consent for their medical data to be analyzed and published in an anonymized format for medical research purposes.

Study Assumption

During public health emergencies, combining on-site and remote clinical trial management models could guarantee the health and treatment needs of clinical trial participants.

Study Design

During the COVID-19 epidemic, the ultimate goal of the management measures was to ensure the treatment of the participants and the implementation of the trials while avoiding the spread of the disease. In this case, remote management was the determinant of strategy effectiveness. Therefore, in addition to the COVID-19 infection rate of the participants, we selected the indicators that were important to the trials and the participants: protocol compliance rate, rate of loss to follow-up, rate of participant withdrawal, rates of disease progression and mortality, and detection rate of monitoring problems.

Clinical trials that were ongoing at the hospital from February 1, 2020, were eligible to be enrolled in the study. The following data were collected through July 31, 2020: numbers of outpatient and inpatient visits, numbers of enrolled participants, visits that were not performed according to protocol, participants who were not administered according to protocol, losses to follow-up, withdrawals, cases of disease progression, and deaths. These data allowed us to assess the clinical trials during the first COVID-19 outbreak from February to April 2020 and during the second outbreak from June to July 2020. The corresponding data during the periods from February to April 2019 and June to July 2019 were also collected. Data were extracted from the hospital information system (HIS) and clinical trial management system (CTMS). We also collected data regarding the use of the remote monitoring system from February 01, 2020, to July 31, 2020, including the frequency of use, number of trials checked, and number of participants checked.

Statistical Methods

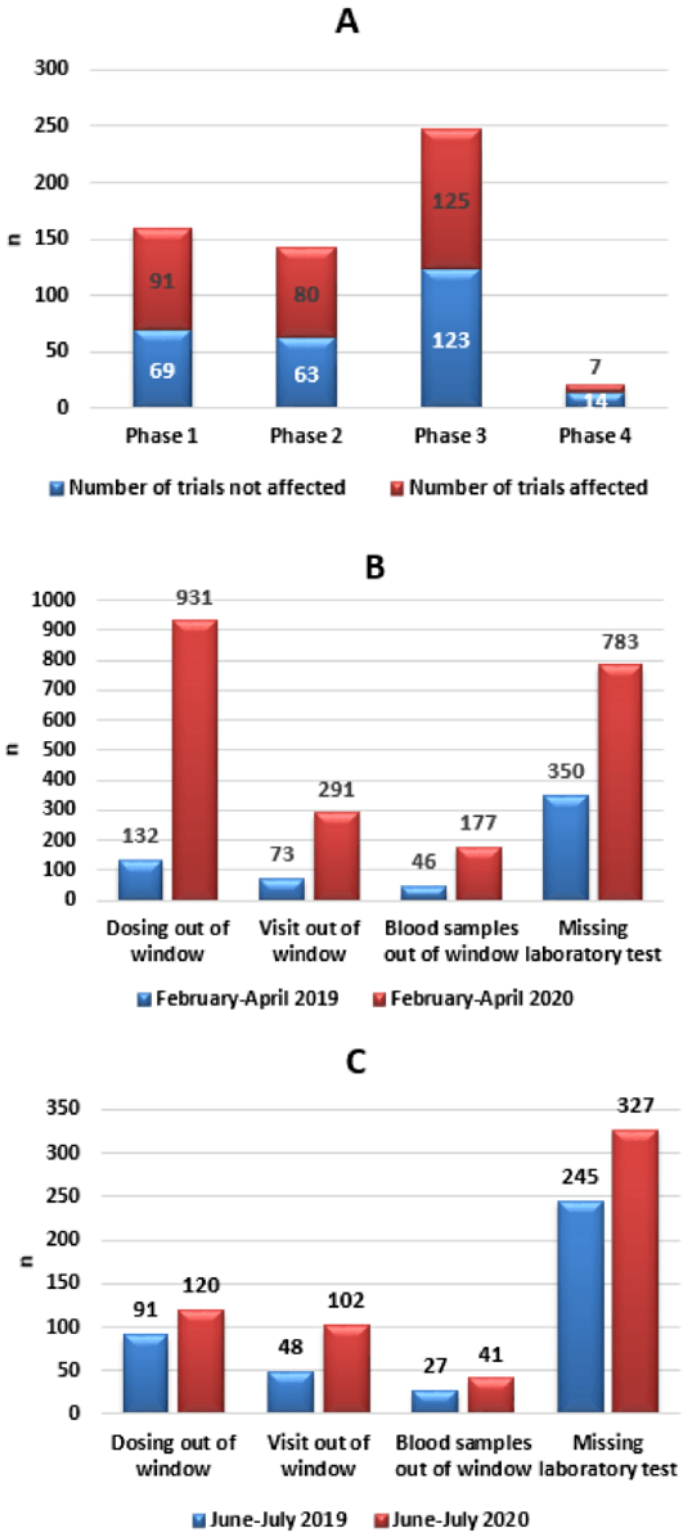
To assess the impact of the COVID-19 epidemic on the conduct of clinical trials at our hospital, we compared the trial data from February to April 2020 and June to July 2020 with data from February to April 2019 and June to July 2019, respectively. To assess the effects of our hospital's modifications to the clinical trials after the first COVID-19 outbreak, we compared the trial data obtained during February and July in 2020. All statistical analyses were performed using SPSS, version 22.0 (IBM Corporation). Categorical data were reported as frequencies or percentages, and differences were assessed for significance using the chi-square test.

Results

Impact of the COVID-19 Outbreak on Clinical Trials at Beijing Cancer Hospital

On July 31, 2020, 572 trials of investigational drugs were being conducted at Beijing Cancer Hospital. More than 50% of all trials (303/572, 53.0%) were affected by the COVID-19 outbreak, including 65 international trials and 238 domestic trials. Of the 303 affected trials, 91 were phase 1 trials, accounting for 56.9% (91/160) of ongoing phase 1 trials. Only 28 new patients were enrolled each month through the two outbreaks, which was 6 times fewer than the monthly enrollment in 2019, and the numbers of various types of protocol deviations were higher than for the same period in 2019. The most frequent deviations were "dosing out of window" and "visit out of window." Due to the outbreak, 42 participants withdrew from trials and 39 were lost to follow-up. The impact of the COVID-19 outbreak on the clinical trials at our hospital is summarized in Figure 1.

Figure 1. The impact of the COVID-19 outbreak on clinical trials at our hospital. (A) Distribution of product phases affected by the outbreak. (B) Protocol violations from February to April 2020 compared to February to April 2019. (C) Protocol violations from June to July 2020 compared to June to July 2019.



Modifications to Clinical Trial Management During the First COVID-19 Outbreak

Remote Trial Approval and Initiation

Starting in February 2020, the hospital implemented information-communication technology to reduce human interactions during clinical trials. The hospital connected its CTMS to an external network so that staff could conduct numerous trial-related activities on the web, including trial review, application for ethics approval, review of collaboration agreements to conduct trials, quality control, and other filings. Trial sponsors and contract research organizations were able to review trials using the same system. Software was also

developed to allow our hospital and principal investigators to initiate new trials if justified.

Remote Trial Monitoring

A remote monitoring system, relying on DPAP technology, integrating various business systems of the hospital, such as the HIS, electronic medical record, laboratory information system, picture archiving and communication system, and radiology information system, was implemented to allow monitors to conduct virtual "site visits" under a virtual private network (VPN). Monitors could check all the medical treatment data of the participants in hospital as authorized and receive a panoramic data view of the participants under the dimensions of medical treatment, examination, medical records, and medical orders. Figure 2 shows the remote monitoring technology roadmap, and Figure 3 shows a screenshot of a panoramic view of a participant's data in the remote monitoring system.

Figure 2. The remote monitoring technology roadmap. CRA: clinical research associate; DPAP: data processing and application platform; EMR: electronic medical record; GAP: gatekeeper; HIS: hospital information system; LIS: laboratory information system; PACS: picture archiving and communication system; RIS: radiology information system; VPN: virtual private network.

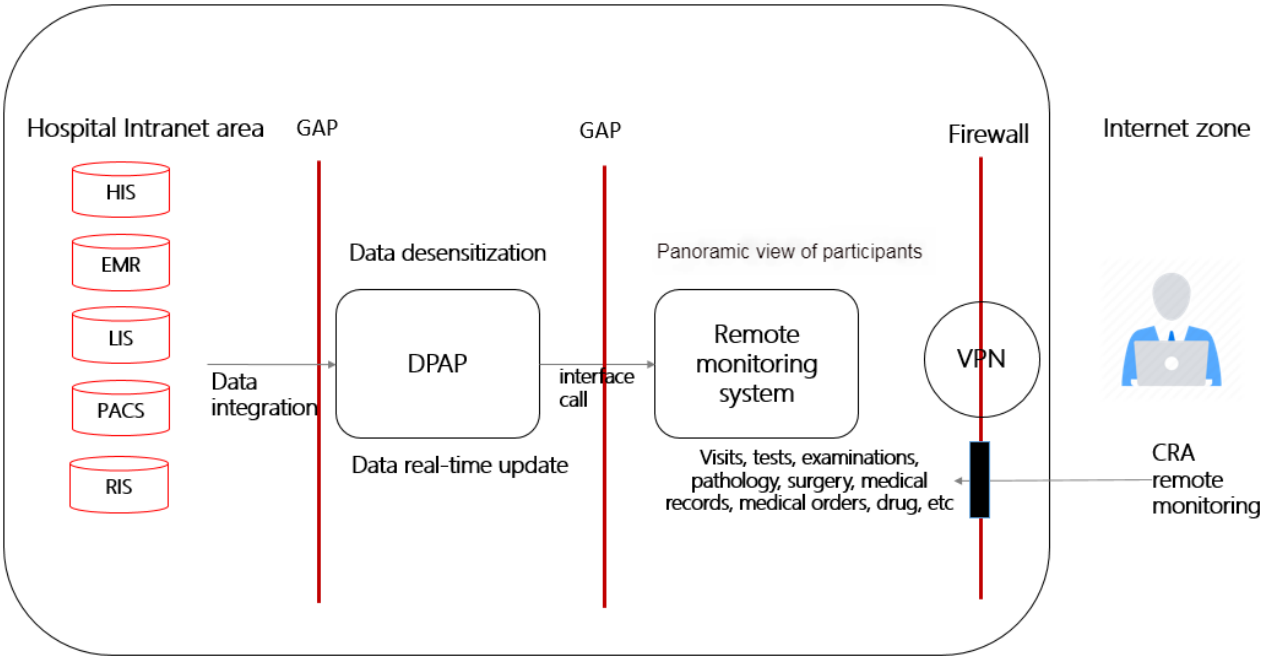


Figure 3. Screenshot of the panoramic view of a clincial trial participant in the remote monitoring system.

Remote Visits and Treatment of Participants

Trial participants who were able to return to hospital upon the resumption of trials were tested for SARS-CoV-2 infection based on nucleic acid detection. Hospitalized participants were examined for COVID-19 symptoms by chest computerized tomography. Beds were placed with sufficient distancing.

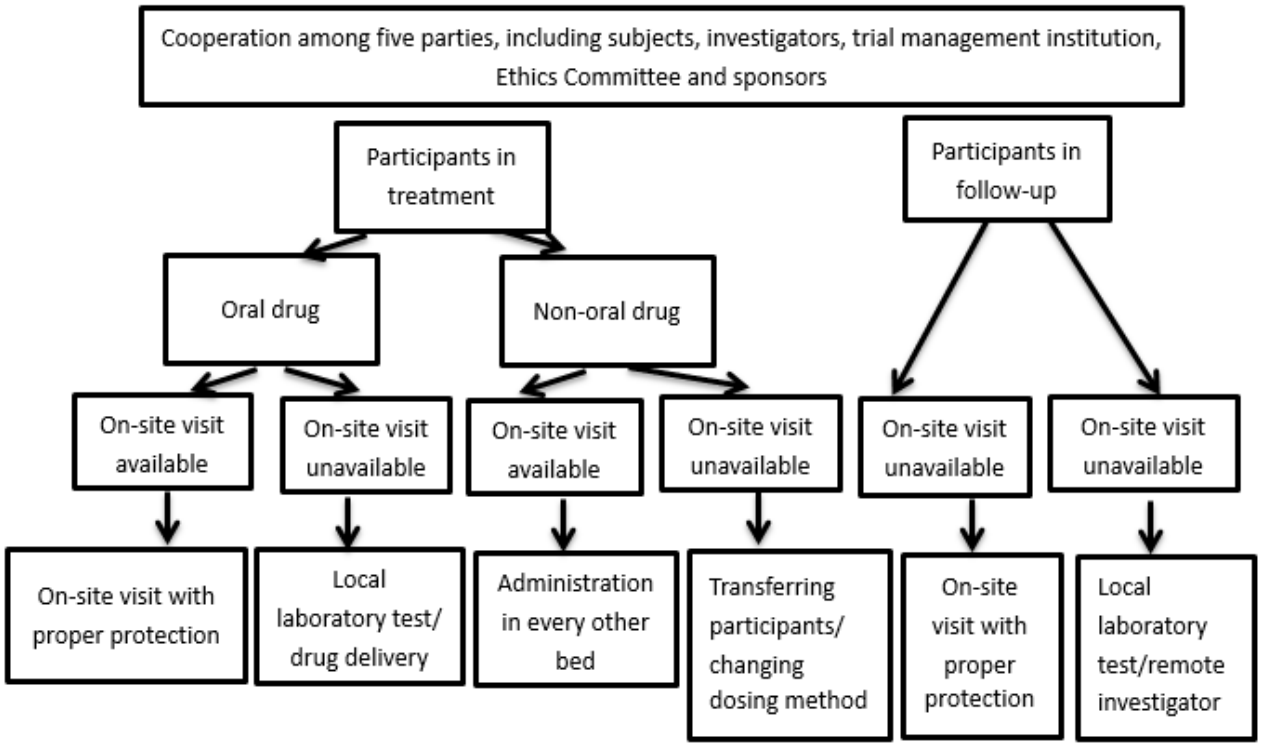
Trial participants who could not return to hospital were "visited remotely": the principal investigator of the trial was located an

appropriate hospital that was relatively close to the participant's residence. Priority was given to hospitals that were also participating in the trial, followed by hospitals that had been certified by the National Medical Products Administration to be following good clinical practice [11]. Trial investigators remotely reviewed examination results through email or fax, and they generated the corresponding medical records, which were stored on a cloud-based platform from July 2020.

Trial medications for oral use were sent to these participants if the trial investigator judged that the drug treatment could continue based on the participant's examination results. Medications were sent by mail using the hospital's standard operating procedures for mailing of investigational products during major public health emergencies. The participants sent back relevant data by mail.

In some cases, trial participants originally scheduled for intravenous drug administration were switched to oral administration of an equivalent medication (eg, etoposide) to allow remote drug treatment. These protocol changes were enacted only after discussion among the trial investigators and ethical approval from the hospital. If the originally scheduled intravenous drug administration could not be switched to oral administration, participants received the trial medications intravenously at an appropriate local hospital (see above), after local staff had been trained by trial personnel from Beijing Cancer Hospital. Participants' medical records were transferred, as necessary, from Beijing Cancer Hospital to the local hospital. Figure 4 depicts the pathway for remote treatment of trial participants.

Figure 4. Schedule of the panoramic view of a clinical trial participant's visits and treatments according to modified clinical trial management procedures implemented from February 01, 2020.



Comparison of Clinical Trial Data Between the First and Second COVID-19 Outbreaks

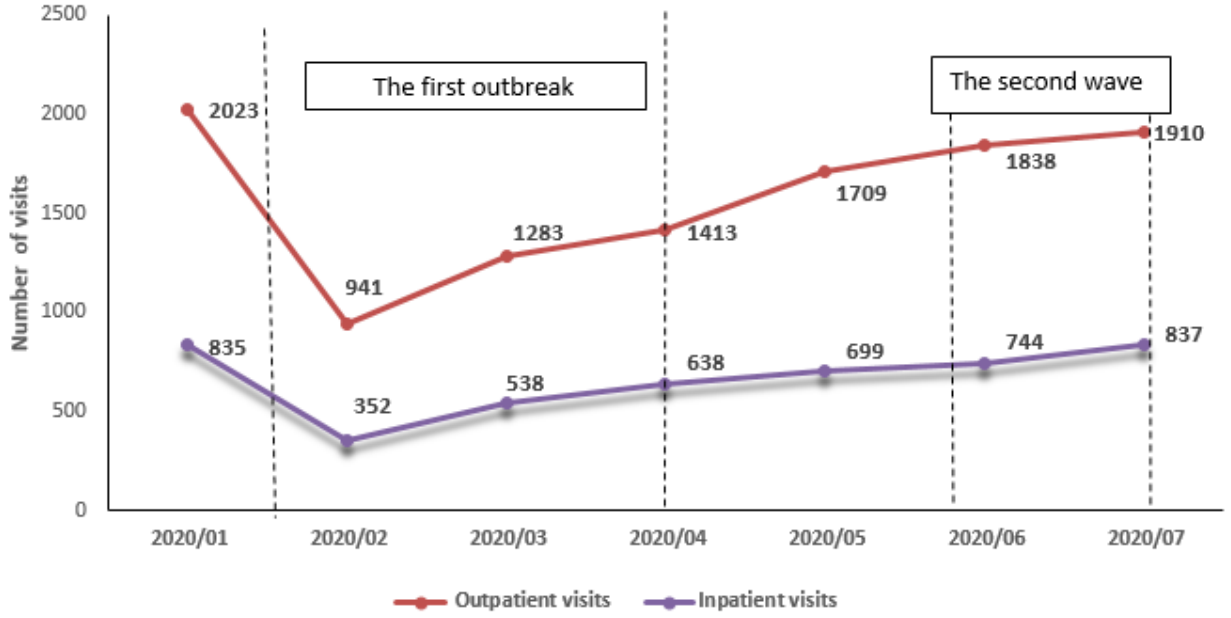
During the first outbreak from February 1 to April 30, 2020, 18 new trials were remotely initiated, and 45 participants were newly enrolled. Moreover, during the second outbreak from June 1 to July 31, 2020, 56 new trials were remotely initiated and 103 participants were newly enrolled, which are nearly double the numbers of trials and enrollments in the first outbreak. By July 31, 572 trials for investigational drugs involving 3718 participants were ongoing at Beijing Cancer Hospital. Between February and July 2020, no infections were recorded among participants or trial staff, and no major procedural errors occurred.

Analysis of the numbers of visits by inpatients and outpatients to Beijing Cancer Hospital for participation in clinical trials showed that both types of visits decreased in February, coinciding with the first COVID-19 outbreak (Table 1). Thereafter, both types of visits continuously increased and had nearly returned to pre-epidemic levels by July (Figure 5).

Table 1. Data on participant visits within clinical trials during the two COVID-19 outbreaks in 2020.

Category	Monthly average for the indicated period	
	First outbreak (February to April)	Second outbreak (June to July)
Return visits to hospital for outpatient examination or treatment	1212	1709
Hospitalizations	509	791
Remote visits	340	109
Drug shipments to participants	281	97
Participants changed from intravenous to oral administration	10	2
Participants transferred to another hospital (times)	16	6
Internet-based diagnostic visits and treatments	0	165
Completed visit/medication (times)	2368	2879
Total required visits/medication (times)	3097	3169

Figure 5. Outpatient and inpatient status of clincial trials at Beijing Cancer Hospital during the COVID-19 epidemic, January 2020 to July 2020.



During the latter part of the first outbreak, 1941/9291 visits with participants (20.89%) were conducted remotely, oral medications were mailed directly to participants, or the participants were treated intravenously at carefully selected local hospitals. The protocol compliance rate was 76.46%. During the second outbreak (June to July), the protocol compliance rate was significantly higher at 90.85% (2879/3169, *P*<.001) (Table 2). Loss to follow-up was significantly smaller during the second outbreak than during the first (7/3718, 0.19% vs 32/3570, 0.89%, *P*<.001), as was participant withdrawal (6/3718, 0.16%, vs 36/3570, 1.00%, *P*<.001). In contrast, the rates of disease progression and mortality did not differ statistically between the two outbreaks (Table 2). Thus, the clinical trials at our hospital remained stable across both COVID-19 outbreaks, and the rate of compliance for the entire period was 85.24% (16007/18778).

During the whole outbreak period, the infection rate of SARS-CoV-2 among personnel involved in clinical trials was 0, and the error rate in the clinical trials was 0.

Table 2. Clinical trial outcomes.

Outcome	Value, n/total (%)		*P* value
	First outbreak	Second outbreak	
Protocol compliance	2368/3097 (76.46)	2879/3169 (90.85)	<.001
Loss to follow-up	32/3570 (0.89)	7/3718 (0.19)	<.001
Withdrawals	36/3570 (1.00)	6/3718 (0.16)	<.001
Disease progression	88/3570 (2.34)	88/3718 (2.37)	.86
Mortality	48/3570 (1.33)	68/3718 (1.83)	.08

Effects of the Measures on Trial Quality During the COVID-19 Epidemic

By July 31, 2020, 176 clinical research associates from 76 sponsors or clinical research organizations had used the remote monitoring system to monitor 1318 participants in 228 trials conducted in 16 departments of Beijing Cancer Hospital. The total number of log-ins to the system was 10,470, and the median number of monitoring visits was 23 (range 1-729). The total number of issues logged in the remote monitoring system was 3820, corresponding to an average of 16.75 per trial and 2.90 per participant. The most frequent findings were errors or omissions on case report forms (950/3820, 24.87% of all issues), lack of compliance with planned trial visits (849/3820, 22.23% of all issues), and lack of compliance with protocol administration (713/3820, 18.66% of all issues). The rate of findings with remote monitoring between original records, informed consent, adverse events, and investigational drugs in 2020 were significantly lower than the rate during the same period in 2019 (P<.001), but the two monitoring approaches were similar in terms of case report forms, concomitant medication, and biological samples (P>.05) (Table 3).

Table 3. Clinical trial–related events detected during on-site monitoring in 2019 and remote monitoring in 2020.[a]

Error or omission in a trial event	Value per capita, n (%)		*P* value
	Remote monitoring in 2020 (n=1318)[b]	On-site monitoring in 2019 (n=1120)[c]	
Case report forms	950 (72.08)	805 (71.88)	.91
Visits not conducted per protocol	849 (64.42)	579 (51.69)	<.001
Drug not dosed per protocol	713 (54.10)	279 (24.91)	<.001
Adverse events	513 (38.92)	643 (57.41)	<.001
Concomitant medications	435 (33.00)	401 (35.80)	.15
Original record	154 (11.68)	874 (78.04)	<.001
Biological samples	146 (11.08)	159 (14.20)	.02
Other drug problems	33 (2.50)	119 (10.63)	<.001
Informed consent	27 (2.05)	119 (10.63)	<.001

[a]Data were compared for the same trials for the period of February to July in 2019 or 2020.

[b]Total errors/emissions=3820; 289.83% cumulative for all events.

[c]Total errors/emissions=3978; 355.18% cumulative for all events.

Discussion

The COVID-19 outbreak is a major global public health emergency, and at its beginning, authorities were taken by surprise [12]. Our analysis of clinical trials at Beijing Cancer Hospital from the beginning of the COVID-19 epidemic showed that 53.0% of investigational drug trials were affected, particularly phase I clinical trials, which involve more extensive interventions. During the first outbreak (February to April 2020), the rate of “dosing out of window” was 6 times higher than the rate during the same period in 2019, while the rate of “visits out of window” was 3 times higher than in 2019. This is especially true of clinical trials of anticancer drugs, which usually require drug administration every 1-2 weeks, and most trials require regular sample collection and patient examinations. These protocol deviations could severely compromise the quality of the trial data as well as the safety and interests of participants [13].

While regulatory authorities in the United States, European Union, and China as well as other associations have guidelines for clinical trial conduct and management during the COVID-19 epidemic [14-17], implementing them in hospitals is not always straightforward. At Beijing Cancer Hospital, discussions were conducted among trial participants, trial investigators, members of the hospital’s Ethics Committee, and trial sponsors to develop modifications to the standard management of the clinical trials and participant visits to protect the health and safety of participants as well as the integrity of the trial. These discussions and the ensuing measures were documented in real time and

archived at the hospital and in the records of the Ethics Committee.

Starting from the middle of the first COVID-19 outbreak, our hospital resumed clinical trials with a series of measures to reduce virus transmission: diagnostic visits and treatment were conducted by appointment only or via the internet; participants were visited remotely; drugs were mailed to participants, including replacement of intravenous treatments with oral treatments that could be mailed; and participants were sent to local hospitals, protecting them from infection during travel to the clinical trial center. To encourage participants to return to the hospital for treatment during the second outbreak, we subjected them to nucleic acid tests and provided beds with adequate separation. Only 50% of beds in each ward were available to patients. These measures may help to explain why more participants returned to the hospital during the second outbreak than during the first. All these measures led to higher protocol compliance rates during the second COVID-19 outbreak than during the first, as well as lower rates of loss to follow-up and withdrawal.

Internet diagnostics and treatment are a new direction in the medical industry, and our experience suggests that this approach can be effective for conducting clinical trials during public health emergencies. Quality control standards are needed to ensure high-quality treatment comparable to that with in-person medicine. To that end, China and other countries have issued regulations and policies to standardize internet diagnostics and treatment [18-20].

Just as internet-based approaches can bring trial clinicians and participants together safely, remote monitoring can allow trial auditors and sponsors to perform necessary reviews. Already in 2019, the US Food Drug and Administration had encouraged sponsors to use remote monitoring for early detection of problems in clinical trials [21], and several national and other agencies have since encouraged remote monitoring in response to the COVID-19 pandemic [14-17]. At Beijing Cancer Hospital, approximately 40% of clinical trials were remotely monitored from February to July 2020, and these activities identified 16.75 issues per trial. Our experience supports the expanding reliance on remote trial monitoring to safeguard patients and trial quality.

Clinical trials in China and abroad face challenges due to the COVID-19 pandemic [22]. Our experience supports the idea that appropriate software and network infrastructure can allow medical facilities to conduct clinical trials effectively [23-25], while properly archiving, managing, and sharing the large amount of data generated [26]. It may even be possible to bypass some aspects of participant follow-up by using smartphones or other wearable devices that can automatically transmit participant data [27]. Conducting a remote clinical trial depends on having a complete information management system, infrastructure for fast and safe delivery of medicines, and a safe, seamless system for managing and transferring trial data.

Remote trial management at Beijing Cancer Hospital can still be improved. For example, we were unable to use the remote informed mode due to technical limitations. Insufficient personnel and delays in data transmission during the epidemic in 2020 contributed to data errors or omissions on case report forms, which accounted for 24.87% of all issues. The system of remote monitoring and automatic collection of clinical trial data should be improved, particularly the extraction of data from paper records.

This study was necessarily retrospective, which increases the risk of selection and information bias. As the pandemic continues, it may be advisable to launch relevant prospective studies to assess the efficacy and efficiency of measures to ensure the smooth conduct of clinical trials during a public health emergency.

Conclusion

Clinical trials have been greatly impacted during the current public health emergency of the COVID-19 pandemic. By using information technology, Beijing Capital Hospital was able to ensure the smooth conduct of hundreds of oncology clinical trials. This success was due to a clinical trial management model combining on-site and remote trial approval, initiation, visits, administration, and monitoring. Our experience provides a reference for clinical trial management under the current pandemic and in future public health emergencies.

Authors' Contributions

ZF and MJ performed data collection and analysis, reviewed the literature, and contributed to drafting the manuscript. KW coordinated and reviewed the manuscript. JL conceptualized and designed the study; supervised, coordinated, and were responsible for the integrity of the data and the accuracy of its analysis; critically reviewed the interpretation of the results; and assisted in the final preparation of the manuscript. All authors read the final manuscript and approved it.

Conflicts of Interest

None declared.

References

1. WHO timeline for COVID-19 response. Webpage in Chinese. World Health Organization. URL: https://www.who.int/zh/news-room/detail/29-06-2020-covidtimeline [accessed 2021-02-24]
2. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://www.who.int/dg/speeches/detail [accessed 2020-04-16]

3. Regmi K, Lwin CM. Impact of non-pharmaceutical interventions for reducing transmission of COVID-19: a systematic review and meta-analysis protocol. BMJ Open 2020 Oct 22;10(10):e041383 [FREE Full text] [doi: 10.1136/bmjopen-2020-041383] [Medline: 33093038]
4. Clinical perspective in the context of COVID-19 — digital health communities in China: interpretation of TIRS literature February (summary). Webpage in Chinese. WeChat. 2020 Feb 14. URL: https://mp.weixin.qq.com/s/QxTqogyqJeGpvPROLFydtw [accessed 2021-02-24]
5. Rubio-San-Simón A, André N, Cefalo MG, Aerts I, Castañeda A, Benezech S, et al. Impact of COVID-19 in paediatric early-phase cancer clinical trials in Europe: a report from the Innovative Therapies for Children with Cancer (ITCC) consortium. Eur J Cancer 2020 Dec;141:82-91 [FREE Full text] [doi: 10.1016/j.ejca.2020.09.024] [Medline: 33129040]
6. Lara Gongoria AB, Fontes Jardim DL, Bastos DA. Oncology clinical trials during the COVID-19 pandemic. Oncology (Williston Park) 2020 Jul 15;34(7):265-269. [Medline: 32674214]
7. Upadhaya S, Yu JX, Oliva C, Hooton M, Hodge J, Hubbard-Lucey VM. Impact of COVID-19 on oncology clinical trials. Nat Rev Drug Discov 2020 Jun;19(6):376-377. [doi: 10.1038/d41573-020-00093-1] [Medline: 32424343]
8. Asaad M, Habibullah NK, Butler CE. The impact of COVID-19 on clinical trials. Ann Surg 2020 Sep 01;272(3):e222-e223 [FREE Full text] [doi: 10.1097/SLA.0000000000004113] [Medline: 32541234]
9. Al-Quteimat OM, Amer AM. The impact of the COVID-19 pandemic on cancer patients. Am J Clin Oncol 2020 Jun;43(6):452-455 [FREE Full text] [doi: 10.1097/COC.0000000000000712] [Medline: 32304435]
10. Liang W, Guan W, Chen R, Wang W, Li J, Xu K, et al. Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. Lancet Oncol 2020 Mar;21(3):335-337 [FREE Full text] [doi: 10.1016/S1470-2045(20)30096-6] [Medline: 32066541]
11. Announcement of the State Food and Drug Administration and the National Health Commission on issuing the quality management standards for drug clinical trials. Webpage in Chinese. National Medical Products Administration, Health Commission of the People's Republic of China. 2020 Apr 27. URL: http://www.nhc.gov.cn/yzygj/s7659/202004/1d5d7ea301f04adba4c4e47d2e92eb96.shtml [accessed 2021-02-24]
12. Ledford H. Coronavirus shuts down trials of drugs for multiple other diseases. Nature 2020 Apr;580(7801):15-16. [doi: 10.1038/d41586-020-00889-6] [Medline: 32214240]
13. Mohan S, Mehra M, Petrizzo M, Katz T. A toolkit for the management of protocol deviations. Ther Innov Regul Sci 2016 Nov;50(6):791-800. [doi: 10.1177/2168479016647987] [Medline: 30231739]
14. FDA guidance on conduct of clinical trials of medical products during the COVID-19 public health emergency. US Food & Drug Administration. 2020 Mar. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/fda-guidance-conduct-clinical-trials-medical-products-during-covid-19-pandemic [accessed 2021-02-24]
15. Brennan S. EMA offers guidance on conducting trials during the COVID-19 pandemic. Regulatory Affairs Professionals Society. 2020 Mar 20. URL: https://www.raps.org/news-and-articles/news-articles/2020/3/ema-offers-guidance-on-conducting-trials-during-th [accessed 2021-02-24]
16. Considerations to support clinical trial monitoring oversight during COVID-19. Association of Clinical Research Organizations. 2020 Mar 13. URL: https://www.acrohealth.org/wp-content/uploads/2020/03/ACRO-Statement-on-Monitoring-Oversight-FINAL-3.13.20.pdf [accessed 2021-02-24]
17. Guidelines for the management of drug clinical trials during COVID-19 outbreaks (trial). Webpage in Chinese. National Medical Products Administration, Health Commission of the People's Republic of China. 2020 Jul 14. URL: https://www.nmpa.gov.cn/xxgk/ggtg/qtggtg/20200715110101939.html [accessed 2021-02-24]
18. A risk-based approach to monitoring of clinical investigations questions and answers. US Food & Drug Administration. 2019 Mar. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/risk-based-approach-monitoring-clinical-investigations-questions-and-answers [accessed 2021-02-24]
19. Reply to suggestions on supporting the construction of "Internet + Education" and "Internet + Medical and Health" demonstration zones. Webpage in Chinese. National Health Commission of the People's Republic of China. 2020 Jul 15. URL: http://www.nhc.gov.cn/wjw/jiany/202007/b2bc9bf348d54c869c25b8eb67c2038f.shtml [accessed 2021-02-24]
20. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
21. Al-Shamsi HO, Alhazzani W, Alhuraiji A, Coomes EA, Chemaly RF, Almuhanna M, et al. A practical approach to the management of cancer patients during the novel coronavirus disease 2019 (COVID-19) pandemic: an international collaborative group. Oncologist 2020 Jun;25(6):e936-e945 [FREE Full text] [doi: 10.1634/theoncologist.2020-0213] [Medline: 32243668]
22. COVID-19 and clinical trials: the Medidata perspective. Medidata. 2020 Jul 15. URL: https://www.medidata.com/en/insight/covid-19-and-clinical-trials-the-medidata-perspective/ [accessed 2021-02-24]
23. Colbert LE, Kouzy R, Abi Jaoude J, Ludmir EB, Taniguchi CM. Cancer research after COVID-19: where do we go from here? Cancer Cell 2020 May 11;37(5):637-638 [FREE Full text] [doi: 10.1016/j.ccell.2020.04.003] [Medline: 32396859]

24. Cerreta F, Ritzhaupt A, Metcalfe T, Askin S, Duarte J, Berntgen M, et al. Digital technologies for medicines: shaping a framework for success. Nat Rev Drug Discov 2020 Sep 12;19(9):573-574. [doi: 10.1038/d41573-020-00080-6] [Medline: 32398879]
25. Li G, Yin C, Zhou Y, Wang T, Chen J, Liu Y, et al. Digitalized adaptation of oncology trials during and after COVID-19. Cancer Cell 2020 Aug 10;38(2):148-149 [FREE Full text] [doi: 10.1016/j.ccell.2020.06.018] [Medline: 32634378]
26. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. Nat Med 2020 Apr;26(4):459-461 [FREE Full text] [doi: 10.1038/s41591-020-0824-5] [Medline: 32284618]
27. Russell C, Ammour N, Wells T, Bonnet N, Kruse M, Tardat A, et al. A pilot study to assess the feasibility of collecting and transmitting clinical trial data with mobile technologies. Digit Biomark 2018;2(3):126-138 [FREE Full text] [doi: 10.1159/000493883] [Medline: 32095763]

Abbreviations

CTMS: clinical trial management system
DPAP: data processing and application platform
HIS: hospital information system
VPN: virtual private network

Edited by G Eysenbach; submitted 27.12.20; peer-reviewed by R Chen, H Bhatt; comments to author 21.01.21; revised version received 26.01.21; accepted 15.02.21; published 02.03.21.

Please cite as:
Fu Z, Jiang M, Wang K, Li J
Minimizing the Impact of the COVID-19 Epidemic on Oncology Clinical Trials: Retrospective Study of Beijing Cancer Hospital
J Med Internet Res 2021;23(3):e26799
URL: https://www.jmir.org/2021/3/e26799
doi:10.2196/26799
PMID:33591924

Viewpoint

Adoption of COVID-19 Contact Tracing Apps: A Balance Between Privacy and Effectiveness

Emily Seto[1,2,3], PhD, PEng; Priyanka Challa[1], MHI; Patrick Ware[2], MPH, PhD

[1]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[2]Centre for Global eHealth Innovation, University Health Network, Toronto, ON, Canada

[3]Techna Institute, University Health Network, Toronto, ON, Canada

Corresponding Author:
Emily Seto, PhD, PEng
Institute of Health Policy, Management and Evaluation
Dalla Lana School of Public Health
University of Toronto
155 College Street, Suite 425
Toronto, ON, M5T 3M6
Canada
Phone: 1 416 669 9295
Email: emily.seto@utoronto.ca

Abstract

With the relative ubiquity of smartphones, contact tracing and exposure notification apps have been looked to as novel methods to help reduce the transmission of COVID-19. Many countries have created apps that lie across a spectrum from privacy-first approaches to those that have very few privacy measures. The level of privacy incorporated into an app is largely based on the societal norms and values of a particular country. Digital health technologies can be highly effective and preserve privacy at the same time, but in the case of contact tracing and exposure notification apps, there is a trade-off between increased privacy measures and the effectiveness of the app. In this article, examples from various countries are used to highlight how characteristics of contract tracing and exposure notification apps contribute to the perceived levels of privacy awarded to citizens and how this impacts an app's effectiveness. We conclude that finding the right balance between privacy and effectiveness, while critical, is challenging because it is highly context-specific.

(J Med Internet Res 2021;23(3):e25726) doi:10.2196/25726

KEYWORDS

mobile apps; COVID-19; contact tracing; exposure notification; privacy; effectiveness; app; surveillance; tracing; transmission; security; digital health

Introduction

Many countries around the world have released contact tracing and exposure notification apps in an attempt to help combat the spread of COVID-19 [1,2]. However, the technologies used, adoption rates, and potential impact of the apps have been extremely varied across countries. Moreover, each country has developed contact tracing apps that meet the level of privacy required for their citizens. Often, increased privacy has been deemed a fair trade-off for a decrease in the potential effectiveness of the app.

Privacy-Related Characteristics of COVID-19 Apps

Although privacy laws provide a foundation that can inform the design and implementation of exposure notification and true contact tracing apps [3], it is the types of technologies used (eg, quick response [QR] codes, GPS, Bluetooth Low Energy) and the way they are applied within those legal frameworks that determine the level of privacy afforded to citizens. The important distinction between contact tracing apps and exposure notification apps is that the former collects tracking data so that public health authorities can determine who individuals have been in contact with, as well as the location and time of the contact. On the other hand, exposure notification apps collect only the data required to determine if an individual may have

been in close contact with someone who has been identified as being positive for COVID-19, which provides significantly more privacy. Table 1 expands upon research by Liu and Guo [4], which was published in the initial months of the pandemic. Specifically, it presents privacy-related characteristics of COVID-19 apps in various countries. The examples have been selected to demonstrate a spectrum of privacy-related features in countries where information about their apps is publicly available, as well as to demonstrate that the optimal balance between privacy and effectiveness may be culturally dependent.

Table 1. Privacy characteristics of COVID-19 contact tracing and exposure notification apps.

Country	App name (month of launch)	Voluntary or mandatory	Technology	App data bound by privacy laws	Consent for data sharing required	Centralized or decentralized data storage	Approximate adoption rate, % (month of reporting)
Australia	COVIDSafe (April 2020) [5]	Voluntary	Bluetooth [5]	Yes	Yes	Centralized	21.6 (July 2020) [6]
China	HealthCode (February 2020) [7]	Voluntary (required to move around cities)	May use GPS or records of individual's location	Yes	No	Information not found	~64 (April 2020)[a] [8]
Canada	COVID Alert (July 2020) [9]	Voluntary	Bluetooth	Yes	Yes	Decentralized	~15 (December 2020)[a,b] [10]
Germany	Corona-Warn-App (June 2020) [11]	Voluntary	Bluetooth	Yes	Yes	Centralized (pseudonymized contact identifiers)	1.4 (July 2020) [6]
Hong Kong	StayHomeSafe (April 2020) [12]	Mandatory for 14-day home quarantine	Bluetooth and geofencing technology using wristbands	Yes	Yes	Decentralized	Information not found
Hong Kong	LeaveHomeSafe (November 2020) [12]	Voluntary	QR[c] codes	Yes	Yes	Decentralized	Information not found
New Zealand	NZ COVID Tracer App (May 2020) [13]	Voluntary	Bluetooth, QR codes, location alert through push notifications	Yes	Yes	Decentralized	10.7 (July 2020) [6]
Russia	Social Monitoring (April 2020) [14]	Mandatory for individuals with COVID-19; voluntary for others	App seeks consent to access Bluetooth, GPS, and camera	Yes	Yes	Centralized	Information not found
Singapore	TraceTogether (March 2020) [15]	Voluntary (required to move around the city)	Bluetooth	Yes	Not for infected persons	Centralized	70 (December 2020) [16]
South Korea	Self-Quarantine Safety Protection (March 2020) [17]	Required for new arrivals for 2 weeks (telephone calls are alternative option)	Bluetooth, GPS, credit card transactions, surveillance cameras, and others	Yes	Yes	Centralized	Information not found
South Korea	Corona 100m (February 2020) [18]	Voluntary	GPS, uses data from public government sources	Yes	No	Decentralized	~2 (end of February 2020, 3 weeks after rollout)[a] [18]
Saudi Arabia	Tawakkalna (May 2020) [19]	Voluntary	GPS	Information not found	Information not found	Centralized	4.9 (July 2020) [6]
United Kingdom	NHS COVID-19 (September 2020) [20]	Voluntary	Bluetooth 4.0 or higher	Yes	Yes	Centralized (nonidentifiable information)	40 (October 2020) [21]

[a]Estimated adoption rate based on number of downloads divided by the country's entire population.

[b]In Canada, the app was rolled out in Ontario first; as of December 2020, only 9 of the 13 provinces and territories had adopted the app.

[c]QR: quick response.

Privacy Measures Related to Adoption

Beyond the inherent value of privacy as a human right and as a protection against bias and stigma, the intent of increasing an app's privacy measures is to broaden its adoption. This has been confirmed by studies conducted since the start of the pandemic, which have demonstrated the important role of trust and perceived privacy in influencing the adoption and use of COVID-19 apps [22-24]. The perception of privacy is not only determined by a user's interpretation of existing safeguards (eg, the underlying technology, whether the app is voluntary, and the degree of data centralization), but also requires those details be effectively communicated to citizens, which is not always the case. For example, an analysis of COVID-19 app privacy policies concluded that improvements to the readability of privacy policies could lead to increased usage [25].

Modelling by a research group at Oxford has indicated that the pandemic can be stopped if about 60% of the population uses the app. Lower adoption rates of 15% would still reduce infection and deaths by about 8% and 6%, respectively [26]. At the time of writing, Singapore and China have been able to reach the 60% threshold, while many other countries have reported adoption rates greater than 15%, including Ireland, Canada (ie, Ontario, where the app was first launched), Germany, and Iceland (the first country in Europe to launch their app) [27].

While adoption rates are affected by numerous nuanced factors, they are likely highly influenced by three broad categories of privacy-related factors highlighted in Table 1. First, apps that are mandatory would presumably result in higher adoption rates than voluntary apps. For example, although use of the apps in Singapore and China is technically voluntary, the apps are required for citizens to move around freely in cities, which may have resulted in the high adoption rates of those apps. Specifically, citizens in China are unable to move freely within cities and enter establishments without showing their color-coded individual QR codes at checkpoints. A color code (green, yellow, or red) is assigned depending on the user's travel history and health status, with green allowing unrestricted movement, and yellow and red indicating different quarantine requirements [28].

Second, the technology employed in the apps will largely dictate how intrusive the apps are to an individual's privacy. Public transparency about the technology used may also increase confidence in the use of the app. To increase adoption, some countries—including Canada, Germany, and the United Kingdom—have opted to use the well-documented and highly vetted Google and Apple Exposure Notification application programming interface (API), which enables the swapping of anonymous identifier beacons (ie, random strings of numbers that are frequently changed) between phones in close proximity via Bluetooth Low Energy [29]. This provides a very high level of privacy because no identifiable data is transmitted. Other countries, such as China, have opted for technology that tracks the location of individuals (eg, GPS), which may be a deterrent to their use. Local Chinese governments have developed their own apps with algorithms that assign the color code, but little information has been made available on the details of how the algorithms work [28]. As another example, in South Korea, all people coming into the country are required to quarantine for two weeks and to download the "Self-Quarantine Safety Protection" app, which tracks a person's movement via GPS to monitor compliance with isolation procedures (telephone calls are an option if someone does not have a smartphone) [30]. After the two-week quarantine period, the app tells the user that they are able to delete the app from their phones. In addition, Corona 100m is a voluntary app in South Korea that was built by a private developer after the government made certain data about patients with COVID-19 freely available [31]. The Corona 100m app shows the location of people infected with the virus, the date the infection was confirmed, and the nationality, sex, and age of the infected person. Alerts are sent to users when they are within 100 meters of the latest tracked location visited by someone positive for COVID-19. Data used by the app comes from smartphone location logs, credit and debit card transactions, and an extensive network of surveillance cameras [32].

Third, data governance in terms of privacy laws related to the app data, user consent for data sharing, and centralization of data storage are also important factors that can impact citizens' comfort level with using the apps. Countries with data governance laws and policies that protect privacy, such as the United Kingdom and Canada, appear to have relatively good adoption rates. On the other hand, a lower adoption rate was reported in Saudi Arabia, where it is not clear if the apps are bound by privacy laws and whether consent for data sharing is required. Some countries have centralized the storage of data into a database controlled by a public health authority, while others have decentralized data storage (ie, data is stored only on an individual's smartphone). While a centralized data storage system could provide added value through the ability to analyze the data for trends and adoption information, decentralized systems may invoke more trust in the app, which could drive up adoption. Germany's exposure notification app was initially developed to support a centralized approach, but was met with much criticism, leading to a change to a decentralized model [33].

The Privacy Versus Effectiveness Trade-Off

The trade-off between privacy and effectiveness is apparent at both the individual and system level. For example, apps leveraging the Google and Apple Exposure Notification API, which was designed with a privacy-first approach, do not provide users with the identity of the person with COVID-19 that they were in close proximity to, or information about the location or time of the potential exposure beyond the fact that it was within the past 14 days. Therefore, the user is provided with little context to determine the actual risk (eg, whether personal protective equipment was used) and how long they should be exercising extra precaution to reduce the risk of transmission of the virus.

The limited collection of data and decentralized systems that are used to protect privacy also hinder the ability of governments

to analyze aggregate data, including demographics, time stamps, and geolocalization, which could inform the design and implementation of more targeted public health strategies. Specifically, COVID-19 apps that prioritize effectiveness over privacy could aggregate reliable demographic information about potentially exposed users that is grounded in a specific place and time. In its absence, governments and public health officials must largely rely on information gathered when patients positive for COVID-19 interact with the health system (which is almost always temporally dissociated from the time of infection) or human-driven contact tracing, which relies on citizens' recall and willingness to report accurate information. While useful, these data may be less reliable and comprehensive compared to data collected via a COVID-19 app, thus limiting the data's ability to inform the types of targeted interventions that could simultaneously decrease the spread of COVID-19 and avoid the negative societal consequences of more generalized lockdowns.

Furthermore, the lack of detailed and centralized data limits evaluations of the effectiveness of these apps. For instance, many exposure notification systems will only be able to determine the number of downloads and the number of users who have chosen to identify themselves as positive for COVID-19 through the app. It is not known whether the user has since deleted the app or has chosen to turn off Bluetooth. Other unknowns include how many people have been notified of a potential exposure, how many people chose to be tested because of an alert, and how many people tested positive for COVID-19 earlier than they would have otherwise due to an alert. In these situations, there is mainly anecdotal effectiveness evidence of users getting alerted about an exposure, getting tested, and then modifying their behavior to reduce the transmission of the virus. Evidence of the effectiveness of the app can help drive adoption of the app, as well as inform future implementations of and improvements to the app. Furthermore, the perception that the app is effective may incentivize individuals to download and use the app, which in itself would presumably increase its effectiveness.

Other Factors Impacting the Adoption of COVID-19 Apps

In addition to privacy concerns, a barrier to adoption is the inability for some to download the app. While the proportion of people who own smartphones is high and increasing (upwards of 80% in some countries), there is still a considerable number of people who do not own smartphones [34]. In addition, a criticism of using the Google and Apple Exposure Notification API is that it works only on phones that were released in the past five years or so, which could have the effect of excluding lower-income communities that may have particularly high rates of COVID-19 transmission [35]. Singapore's innovative solution to help reach citizens that are unable to download the TraceTogether app was to distribute a device called the TraceTogether token, which works by swapping identifier beacons via Bluetooth, similar to the app [36].

There are other methods that have been used to try to increase adoption of voluntary exposure notification apps. One is to incorporate features in the app to increase its perceived value [37]. For example, some exposure notification apps can provide COVID-19 test results. In the United Kingdom, the NHS COVID-19 app has features such as ordering COVID-19 tests, receiving test results, regional risk score alerts, symptom recorders, and a self-isolation countdown and advice [38].

A second way to increase adoption of these apps is through social influence and media campaigns. In Canada, influential brands and high-profile individuals like athletes have partnered to promote the use of the COVID Alert app [39]. New Zealand has used humor and creativity in their efforts to inform citizens about the NZ COVID Tracer app, and the pandemic more generally, through comedic skits with well-known television personalities and a call for filmmakers to submit short videos [40]. Opportunities to communicate the privacy safeguards as well as personal and societal benefits should also be explored in these campaigns.

Finally, reducing user effort by making the app easier to download and use may also increase adoption. For example, during the launch of the COVID Alert app in Ontario, Canada, a government alert was sent to smartphones regarding the app, with information on how to download the app, which runs in the background after setup without further user interaction.

Conclusion

With increased privacy, there are inherent trade-offs in the effectiveness of COVID-19 contact tracing and exposure notification apps. The effectiveness of the apps might be impossible to evaluate fully due to the lack of collected data, especially for apps with privacy-first approaches, as well as confounding factors like community lockdowns. However, given the assumption that higher adoption translates into increased effectiveness, broadening adoption of voluntary apps is a goal of many countries, which can be achieved through several techniques. These include investing in a promotional campaign that may involve hiring a professional marketing firm, partnering with high-profile personalities to endorse the app, and increasing ease of app download via smartphone alerts by the government that link to the app. While the level of privacy required for a COVID-19 contact tracing and exposure notification app will depend on factors including whether it is voluntary, the underlying technology, and degree of data centralization, translation of those important safeguards into a user's perception of privacy will occur within the context of the norms and values of their country. Therefore, striking the right balance between privacy and effectiveness requires careful consideration, especially as the urgency to reduce transmission of the virus evolves based on fluctuating case numbers and vaccination efforts.

Conflicts of Interest

None declared.

References

1. Sharma T, Bashir M. Use of apps in the COVID-19 response and the loss of privacy protection. Nat Med 2020 Aug 26;26(8):1165-1167. [doi: 10.1038/s41591-020-0928-y] [Medline: 32457443]
2. Abrahams N, Flockhart F, Cramer S, Cwalina C, Evans M, Gamvros A, et al. Contact tracing apps: A new world for data privacy. Norton Rose Fullbright. 2020. URL: https://www.nortonrosefulbright.com/en-us/knowledge/publications/d7a9a296/contact-tracing-apps-a-new-world-for-data-privacy#Germany [accessed 2020-11-01]
3. Bradford L, Aboy M, Liddell K. COVID-19 Contact Tracing Apps: A Stress Test for Privacy, the GDPR and Data Protection Regimes. Journal of Law and the Biosciences 2020 Jun;7(1):3. [doi: 10.1093/jlb/lsaa034] [Medline: 32728470]
4. Li J, Guo X. Global Deployment Mappings and Challenges of Contact-tracing Apps for COVID-19. SSRN 2020 May 26:1-7. [doi: 10.2139/ssrn.3609516]
5. Abrahams N, Cwalina C, Evans M, Flockhart F, Gamvros A, Lennon J. Contact tracing apps in Australia. Norton Rose Fulbright. 2020 Dec 01. URL: https://www.nortonrosefulbright.com/-/media/files/nrf/nrfweb/contact-tracing/australia-contact-tracing.pdf?revision= 9f35a88a-4124-4c48-b38f-68e86a187050&la=en-au [accessed 2020-12-22]
6. Chan S. COVID-19 Contact Tracing Apps Reach 9% Adoption In Most Populous Countries. SensorTower. 2020 Jul 14. URL: https://sensortower.com/blog/contact-tracing-app-adoption [accessed 2020-12-22]
7. Gamvros A, Cwalina C, Evans M, Flockhart F. Contact tracing apps in China. Norton Rose Fulbright. 2020 May 11. URL: https://www.nortonrosefulbright.com/-/media/files/nrf/nrfweb/contact-tracing/china-contact-tracing.pdf?revision=249d55f4-eb9a-49dd-8491-b8c9c7626691&la=en-au [accessed 2021-03-02]
8. Associated Press. Green symbol on Chinese smartphone screens allows subway travel, hotel check-in; new 'health code' a tool in getting economy on its feet. MarketWatch. 2020 Apr 06. URL: https://www.marketwatch.com/story/green-symbol-health-code-on-smartphone-screens-is-part-of-chinas-bid-to-get-economy-moving-again-2020-04-06 [accessed 2020-12-20]
9. Cassell J, Himo J, Cwalina C, Evans M, Flockhart F, Gamvros A. Contact tracing apps in Canada. Norton Rose Fulbright. 2020 Jul 01. URL: https://www.nortonrosefulbright.com/-/media/files/nrf/nrfweb/contact-tracing/canada-contact-tracing.pdf?revision=fd135992-df8b-45d9-922e-b7b7fbc33f49&la=en-au [accessed 2020-12-20]
10. Download the COVID Alert mobile app to protect yourself and your community. Ontario Government. 2021 Feb 25. URL: https://covid-19.ontario.ca/covidalert [accessed 2020-12-20]
11. Ritzer C, Cwalina C, Evans M, Flockhart F, Gamvros A. Contact tracing apps in Germany. Norton Rose Fulbright. 2020 Jun 23. URL: https://www.nortonrosefulbright.com/-/media/files/nrf/nrfweb/contact-tracing/germany-contact-tracing.pdf?revision=ed379c1d-011d-4cd8-8e65-02c2664e2ba9 [accessed 2020-12-20]
12. Gamvros A, Cwalina C, Evans M, Flockhart F. Contact tracing apps in Hong Kong. Norton Rose Fulbright. 2020 Nov 16. URL: https://www.nortonrosefulbright.com/-/media/files/nrf/nrfweb/contact-tracing/hk-contact-tracing.pdf?revision=de7dc38b-2ce3-4d4e-81a1-59642fa15b26&la=en-cn [accessed 2020-12-20]
13. Getting started with NZ COVID Tracer. New Zealand Ministry of Health. 2020 Dec 21. URL: https://www.health.govt.nz/our-work/diseases-and-conditions/covid-19-novel-coronavirus/covid-19-resources-and-tools/nz-covid-tracer-app/getting-started-nz-covid-tracer [accessed 2020-12-23]
14. Kozyreva A, Cwalina C, Evans M, Flockhart F, Gamvros A. Contact tracing apps in Russia. Norton Rose Fulbright. 2020 May 15. URL: https://www.nortonrosefulbright.com/-/media/files/nrf/nrfweb/contact-tracing/russia-contact-tracing.pdf?revision=801f066d-714b-4148-81ea-d6febc766899&la=en-cn [accessed 2020-12-20]
15. Goggin G. COVID-19 apps in Singapore and Australia: reimagining healthy nations with digital technology. Media International Australia 2020 Aug 14;177(1):61-75. [doi: 10.1177/1329878X20949770]
16. Meah N. TraceTogether adoption rate surpasses 70%, more distribution points to reopen from January 2021. Today. 2020 Dec 23. URL: https://www.todayonline.com/singapore/tracetogether-adoption-rate-surpasses-70-more-distribution-points-reopen-january-2021 [accessed 2020-12-23]
17. Joo J, Shin MM. Resolving the tension between full utilization of contact tracing app services and user stress as an effort to control the COVID-19 pandemic. Serv Bus 2020 Sep 01;14(4):461-478. [doi: 10.1007/s11628-020-00424-7]
18. Huang Y, Sun M, Sui Y. How digital contact tracing slowed COVID-19 in East Asia. Harvard Business Review. 2020 Apr 15. URL: https://hbr.org/2020/04/how-digital-contact-tracing-slowed-covid-19-in-east-asia [accessed 2020-12-19]
19. Khalid T. Coronavirus: How to use Saudi Arabia's 'Tawakkalna' app to get movement permits. Al Arabiya. 2020 May 26. URL: https://english.alarabiya.net/en/coronavirus/2020/05/26/Coronavirus-How-to-use-Saudi-Arabia-s-Tawakkalna-app-to-get-movement-permits [accessed 2020-12-19]
20. White L, Cwalina C, Evans M, Flockhart F, Gamvros A. Contact tracing apps in the UK. Norton Rose Fulbright. 2020 Oct 05. URL: https://www.nortonrosefulbright.com/-/media/files/nrf/nrfweb/contact-tracing/uk-contact-tracing.pdf?la=en-us&revision= [accessed 2020-12-20]

21. Kent C. A comedy of errors: the UK's contact-tracing apps. Medical Device Network. 2020 Oct 06. URL: https://www.medicaldevice-network.com/features/uk-contact-tracing-app-problems/ [accessed 2020-12-21]
22. Kaspar K. Motivations for Social Distancing and App Use as Complementary Measures to Combat the COVID-19 Pandemic: Quantitative Survey Study. J Med Internet Res 2020 Aug 27;22(8):e21613 [FREE Full text] [doi: 10.2196/21613] [Medline: 32759100]
23. Walrave M, Waeterloos C, Ponnet K. Adoption of a Contact Tracing App for Containing COVID-19: A Health Belief Model Approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
24. Abuhammad S, Khabour O, Alzoubi K. COVID-19 Contact-Tracing Technology: Acceptability and Ethical Issues of Use. PPA 2020 Sep 18;Volume 14:1639-1647. [doi: 10.2147/ppa.s276183]
25. Zhang M, Chow A, Smith H. COVID-19 Contact-Tracing Apps: Analysis of the Readability of Privacy Policies. J Med Internet Res 2020 Dec 03;22(12):e21572 [FREE Full text] [doi: 10.2196/21572] [Medline: 33170798]
26. Abueg M, Hinch R, Wu N, Liu L, Probert W, Wu A, et al. Modeling the combined effect of digital exposure notification and non-pharmaceutical interventions on the COVID-19 epidemic in Washington state. medRxiv. Preprint published online on September 2, 2020. [doi: 10.1101/2020.08.29.20184135]
27. Farronato C, Iansiti M, Bartosiak M, Denicolai S, Ferretti L, Fontana R. How to get people to actually use contact-tracing apps. Harvard Business Review. 2020 Jul 15. URL: https://hbr.org/2020/07/how-to-get-people-to-actually-use-contact-tracing-apps [accessed 2020-12-21]
28. Utzerath J, Bird R, Cheng G, Ohara J. Contact tracing apps in China, Hong Kong, Singapore, Japan, and South Korea. Freshfields Bruckhaus Deringer. 2020 Apr 24. URL: https://digital.freshfields.com/post/102g5my/contact-tracing-apps-in-china-hong-kong-singapore-japan-and-south-korea [accessed 2020-12-19]
29. Exposure Notifications: Using technology to help public health authorities fight COVID-19. Google. URL: https://www.google.com/covid19/exposurenotifications/ [accessed 2020-12-12]
30. Smith J, Hyonhee S, Cha S. Ahead of the curve: South Korea's evolving strategy to prevent a coronavirus resurgence. Reuters. 2020 Apr 15. URL: https://www.reuters.com/article/us-health-coronavirus-southkorea-respons-idUSKCN21X0MO [accessed 2020-12-16]
31. Klatt K. Corona apps: South Korea and the dark side of digital tracking. The Brussels Times. 2020 Apr 30. URL: https://www.brusselstimes.com/opinion/108594/corona-apps-south-korea-and-the-dark-side-of-digital-tracking/ [accessed 2020-12-15]
32. Review of Mobile Application Technology to Enhance Contact Tracing Capacity for COVID-19. Johns Hopkins Center for Health Security. 2020 Apr 08. URL: https://www.centerforhealthsecurity.org/resources/COVID-19/COVID-19-fact-sheets/200408-contact-tracing-factsheet.pdf [accessed 2020-12-10]
33. Busvine D, Rinke A. Germany flips to Apple-Google approach on smartphone contact tracing. Reuters. 2020 Apr 26. URL: https://www.reuters.com/article/us-health-coronavirus-europe-tech-idUSKCN22807J [accessed 2020-12-21]
34. Taylor K, Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. 2019 Feb 05. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2020-12-19]
35. Help for COVID Alert. Government of Canada. URL: https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19/covid-alert/help.html [accessed 2020-12-19]
36. TraceTogether, safer together. Singapore Government. 2020 Jul 20. URL: https://www.tracetogether.gov.sg/ [accessed 2020-12-18]
37. Walrave M, Waeterloos C, Ponnet K. Adoption of a Contact Tracing App for Containing COVID-19: A Health Belief Model Approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
38. NHS COVID-19: the new contact-tracing app from the NHS. National Cyber Security Centre. 2020 May 04. URL: https://www.ncsc.gov.uk/information/nhs-covid-19-app-explainer [accessed 2020-12-18]
39. Millions Across Canada Now Using Made-in-Ontario COVID Alert App. Ontario Government. 2020 Oct 15. URL: https://news.ontario.ca/en/release/58831/millions-across-canada-now-using-made-in-ontario-covid-alert-app [accessed 2020-12-19]
40. Bernacki E. Covid19 as wicked problem-solving: did your country change behaviours and make people laugh? LinkedIn. 2020 Jul 08. URL: https://www.linkedin.com/pulse/covid19-wicked-problem-solving-did-your-country-change-ed-bernacki/ [accessed 2020-12-19]

Abbreviations

API: application programming interface
QR: quick response

Edited by G Eysenbach, C Basch; submitted 12.11.20; peer-reviewed by J Li, J Ropero; comments to author 04.12.20; revised version received 24.12.20; accepted 19.02.21; published 04.03.21.

Please cite as:
Seto E, Challa P, Ware P
Adoption of COVID-19 Contact Tracing Apps: A Balance Between Privacy and Effectiveness
J Med Internet Res 2021;23(3):e25726
URL: https://www.jmir.org/2021/3/e25726
doi:10.2196/25726
PMID:33617459

Original Paper

Modeling Predictive Age-Dependent and Age-Independent Symptoms and Comorbidities of Patients Seeking Treatment for COVID-19: Model Development and Validation Study

Yingxiang Huang[1], PhD; Dina Radenkovic[2,3], MD; Kevin Perez[1], MSc; Kari Nadeau[4], MD, PhD; Eric Verdin[1], PhD; David Furman[1], PhD

[1]Buck Institute for Research on Aging, Novato, CA, United States

[2]Guy's & St Thomas' NHS Foundation Trust, London, United Kingdom

[3]King's College London, London, United Kingdom

[4]Division of Pulmonary, Allergy and Critical Care Medicine, Sean N Parker Center for Allergy and Asthma Research, Stanford University, Stanford, CA, United States

Corresponding Author:
David Furman, PhD
Buck Institute for Research on Aging
8001 Redwood Blvd
Novato, CA, 94945
United States
Phone: 1 (415) 209 2000
Email: DFurman@buckinstitute.org

Abstract

Background: The COVID-19 pandemic continues to ravage and burden hospitals around the world. The epidemic started in Wuhan, China, and was subsequently recognized by the World Health Organization as an international public health emergency and declared a pandemic in March 2020. Since then, the disruptions caused by the COVID-19 pandemic have had an unparalleled effect on all aspects of life.

Objective: With increasing total hospitalization and intensive care unit admissions, a better understanding of features related to patients with COVID-19 could help health care workers stratify patients based on the risk of developing a more severe case of COVID-19. Using predictive models, we strive to select the features that are most associated with more severe cases of COVID-19.

Methods: Over 3 million participants reported their potential symptoms of COVID-19, along with their comorbidities and demographic information, on a smartphone-based app. Using data from the >10,000 individuals who indicated that they had tested positive for COVID-19 in the United Kingdom, we leveraged the Elastic Net regularized binary classifier to derive the predictors that are most correlated with users having a severe enough case of COVID-19 to seek treatment in a hospital setting. We then analyzed such features in relation to age and other demographics and their longitudinal trend.

Results: The most predictive features found include fever, use of immunosuppressant medication, use of a mobility aid, shortness of breath, and severe fatigue. Such features are age-related, and some are disproportionally high in minority populations.

Conclusions: Predictors selected from the predictive models can be used to stratify patients into groups based on how much medical attention they are expected to require. This could help health care workers devote valuable resources to prevent the escalation of the disease in vulnerable populations.

(J Med Internet Res 2021;23(3):e25696) doi:10.2196/25696

KEYWORDS

clinical informatics; predictive modeling; COVID-19; app; model; prediction; symptom; informatics; age; morbidity; hospital

Introduction

The COVID-19 pandemic, caused by SARS-CoV-2, continues to burden medical institutions around the world by increasing total hospitalization and intensive care unit (ICU) admissions [1-9]. A better understanding of symptoms, comorbidities, and medications used for pre-existing conditions of patients with COVID-19 could help health care workers identify patients at increased risk of developing a more severe case of the disease [10,11]. Here, we have used self-reported data (symptoms, medications, and comorbidities) from 11,194 users of the COVID-19 Symptom Tracker app [12] to identify previously reported and novel features predictive of patients being admitted to a hospital. Despite the previously reported association between age and more severe disease phenotypes [13-18], we found that a patient's age, sex, and ethnic group were minimally predictive when compared to the patient's symptoms and comorbidities. The most important variables selected by our predictive models were fever, the use of immunosuppressant medication, the use of a mobility aid, shortness of breath, and fatigue. It is anticipated that early administration of preventative treatment in patients with COVID-19 who exhibit a high risk of a severe disease signature may prevent disease progression. Therefore, the variables found could provide clinical decision support.

Methods

Cohort

The COVID-19 Symptom Tracker is a smartphone app to which individuals from the United Kingdom and United States can submit their symptoms daily [19-22]. A total of 3,643,842 users had signed up for the app as of November 1, 2020 [12]. A user can make entries in the app and record features such as symptoms, comorbidities, medications taken for pre-existing conditions, and demographics. Users can log entries whether they have tested positive for COVID-19 or not. They can also create multiple entries per day or one entry every few days. Data from March 24, 2020, to November 1, 2020, were accessed from The Health Data Research Hub for Respiratory Health, in partnership with Secure Anonymised Information Linkage (SAIL) Databank. There were users from both the United States and the United Kingdom, but we only used the UK population to maintain the homogeneity of the data set (Figure S1 in Multimedia Appendix 1). The US population was small and could have a different disease expression than that caused by the virus strain in the United Kingdom. Therefore, we excluded the small US population.

For the study cohort, we extracted all users who tested positive for COVID-19 (n=11,194). Of those users who were positive for COVID-19, some cases were severe enough to require seeking treatment at a hospital (hereinafter referred to as seeking treatment) while others managed their disease at home (Figure S1 in Multimedia Appendix 1). The features we used in all subsequent models are listed in Tables 1-4. All features were binary except for age and BMI, which were continuous, and shortness of breath, fatigue, race, and gender, which were categorical.

Table 1. Features of symptoms investigated in relation to whether a user was seeking treatment. All features were binary except for shortness of breath and fatigue, which were categorical.

Symptoms	Managed at home (N=10,636), n (%)	Admitted into hospital setting (N=558), n (%)
Fever		
1	2761 (26.4)	155 (29.0)
N/A[a]	42 (0.40)	80 (15.0)
Persistent cough		
1	5429 (52.1)	255 (47.7)
N/A	0 (0.0)	80 (15.0)
Diarrhea		
1	2944 (28.3)	178 (33.3)
N/A	0 (0.0)	80 (15.0)
Delirium		
1	1888 (18.1)	137 (25.6)
N/A	0 (0.0)	80 (15.0)
Skipped meals		
1	3967 (38.1)	218 (40.7)
N/A	0 (0.0)	80 (15.0)
Abdominal pain		
1	2452 (23.5)	129 (24.1)
N/A	116 (1.11)	83 (15.5)
Chest pain		
1	2452 (23.5)	129 (24.1)
N/A	116 (1.1)	83 (15.5)
Loss of smell		
1	6251 (60.0)	217 (40.6)
N/A	95 (0.9)	83 (15.5)
Headache		
1	4326 (41.5)	250 (43.0)
N/A	95 (0.9)	83 (15.5)
Sore throat		
1	3862 (37.1)	158 (29.5)
N/A	344 (3.30)	94 (17.6)
Unusual muscle pains		
1	3442 (33.1)	168 (31.4)
N/A	1215 (11.7)	116 (21.7)
Shortness of breath		
N/A	0 (0.0)	80 (15.0)
No	6771 (65.0)	219 (40.9)
Mild	2945 (28.3)	156 (29.2)
Significant	626 (6.01)	65 (12.1)
Severe	71 (0.7)	15 (2.8)
Fatigue		
N/A	0 (0.0)	80 (15.0)

Symptoms	Managed at home (N=10,636), n (%)	Admitted into hospital setting (N=558), n (%)
No	5922 (56.9)	210 (39.3)
Mild	3712 (35.6)	178 (33.3)
Severe	779 (7.5)	67 (12.5)

[a]N/A: data not available or missing.

Table 2. Features of comorbidities investigated in relation to whether a user was seeking treatment. All features were binary.

Comorbidities	Managed at home (N=10,636), n (%)	Admitted into hospital setting (N=558), n (%)
Has diabetes		
1	370 (3.6)	48 (9.0)
N/A[a]	26 (0.2)	1 (0.2)
Has heart disease		
1	209 (2.0)	40 (7.5)
N/A	26 (0.2)	1 (0.2)
Has lung disease		
1	1446 (13.9)	122 (22.8)
N/A	26 (0.2)	1 (0.2)
Is a smoker		
1	294 (2.8)	21 (3.9)
NA	5241 (50.3)	265 (49.5)
Undergoing chemotherapy		
1	29 (0.3)	12 (2.2)
N/A	5184 (49.8)	244 (45.6)
Has kidney disease		
1	91 (0.9)	19 (3.6)
N/A	26 (0.2)	1 (0.2)
Housebound		
1	538 (5.2)	102 (19.1)
N/A	19 (0.2)	0 (0.0)
Uses a mobility aid		
1	183 (1.8)	77 (14.4)
N/A	19 (0.2)	0 (0.0)
Limited activity		
1	886 (8.5)	137 (25.6)
N/A	26 (0.2)	1 (0.2)

[a]N/A: data not available or missing.

Table 3. Features of demographics investigated in relation to whether a user was seeking treatment. Age and BMI were continuous features, while race and gender were categorical.

Demographics	Managed at home (N=10,636)	Admitted into hospital setting (N=558)
Gender, n (%)		
Female	2981 (28.4)	219 (40.9)
Male	7455 (71.6)	316 (59.1)
Race, n (%)		
UK Asian	519 (5.0)	29 (5.4)
UK Black	116 (1.1)	9 (1.7)
UK mixed, White/Black	56 (0.5)	1 (0.2)
UK mixed, other	116 (1.1)	4 (0.7)
UK White	9346 (89.7)	472 (88.2)
UK Chinese	43 (0.4)	4 (0.7)
UK Middle Eastern	89 (0.8)	8 (1.5)
Other	97 (0.9)	5 (0.9)
Prefer not to say	33 (0.3)	3 (0.6)
Age (years), mean (SD)	40.2 (13.6)	47.8 (18.8)
BMI, mean (SD)	26.3 (6.9)	27.9 (8.2)

Table 4. Features of medication history investigated in relation to whether a user was seeking treatment. All features were binary.

Medication	Managed at home (N=10,636), n (%)	Admitted into hospital setting (N=558), n (%)
Takes corticosteroids		
1	719 (6.9)	62 (11.6)
N/A[a]	26 (0.2)	1 (0.2)
Takes immunosuppressants		
1	287 (2.8)	65 (12.1)
N/A	26 (0.2)	1 (0.2)
Takes any blood pressure medications		
1	721 (6.9)	93 (17.4)
N/A	1022 (9.8)	54 (10.1)

[a]N/A: data not available or missing.

Data Processing

We used comorbidities, demographics, and symptoms to predict whether a user would seek treatment using predictive models. For data preprocessing, we first divided the patients with COVID-19 into two groups: (A) negative for seeking treatment, including patients who were positive for COVID-19 who were strictly at home, without ever having to be admitted to a hospital setting (n=10,636) and (B) positive for seeking treatment, including users who were positive for COVID-19 who reported being in a hospital setting (n=558). The average age of group A was 40.2 (SD 13.6) years compared to 47.8 (SD 18.8) years for group B. For group A, we used comorbidities, demographics, and symptoms recorded in the patient's last entry, and for group B, we used the same features as recorded in the patient's last entry prior to the entry where the patient indicates seeking treatment (scenario 1). This means that for users who sought treatment, we used the time point right before a user indicated seeking treatment in a hospital setting and the features at that time point were used for analysis. For users who were always at home, we used the last time point and the features at that time point for analysis (Figure S2, top panel, in Multimedia Appendix 1). In what we call scenario 2, for users who were seeking treatment, if a user indicated that he/she had a feature in any of his/her entries before the day of seeking treatment, we labeled that feature as positive for that user. For users who were always at home, if he/she had a given feature in any entry, we labeled that feature as positive for that user (Figure S2, bottom panel, in Multimedia Appendix 1). Such processing only applied to symptoms since they can change daily, but not to comorbidities, pre-existing medication use, or demographic information.

Imputations

Patients often neglect to report all available fields, so we used the multiple imputations method to impute missing values, a standard procedure to predict missing data using all other features (besides the outcome) that are not missing [23-25]. Multiple imputations were used to impute missing values. Instead of imputing the missing value with a single value, multiple imputations repeatedly sample the data *n* times and impute the missing values *n* times. Different types of data require different imputation methods. We used predictive mean matching for continuous numerical data (age, BMI), polytomous regression for unordered categorical data (gender, race), proportional odds model for ordered categorical data (fatigue, shortness of breath), and logistic regression for binary data (all other features). The predictive variables for the different methods would be all other independent variables, while the outcome would be the missing variable. The most stringent process would only impute the training set, but there were not enough complete instances to have both positive and negative cases; therefore, we imputed the training and testing sets together. Multiple imputations produce *n* imputations, and we pooled *n* imputed matrices together to form a larger training set. Some variables had a large percentage of missing values, as seen in Figure S3, bottom panel (Multimedia Appendix 1). Such variables produced wide ranges of distribution from one imputation to another that are too different from the original distribution (Figure S3, top panel, in Multimedia Appendix 1). Therefore, those variables were removed from the data sets. The final set of variables are all present in Tables 1-4.

Imbalanced Classification

Class imbalance was a problem in our data set, where the minority class (individuals who sought treatment) constituted about 5% of the total COVID-19 population. Upon examining the distribution modality of our data set and discovering that the minority class is multimodal, we decided to use sampling methods to balance the data set and a binary classifier in the next section [26]. The sampling method we employed randomly undersamples the majority class and oversamples the minority class until the data is balanced and the total number of instances is equal to the original number of instances. Such a sampling method was applied in the following predictive model.

Predictive Modeling

We performed an Elastic Net regularized binary classification to select for the most important features in predicting users who needed to seek medical treatment. The data set was divided into 10-fold cross-validation with 10 training and test sets (with a ratio of 70:30). In each of the cross-validations, the resampling method was applied to the training set. Parameters were then tuned for the Elastic Net classifier using the training set, producing the best predictive performance and the most parsimonious number of features. Two parameters can be tuned in Elastic Net, alpha (α) and lambda (λ). Alpha is the mixing parameter, indicating how much least absolute shrinkage and selection operator (LASSO) and Ridge regularization should contribute to the model, while lambda is the amount of shrinkage or regularization the model should apply as a whole. A series of alpha is used in each cross-validation in a grid search with values ranging from 0.0 to 1.0 in steps of 0.05. The alpha that produces the highest area under the receiver operating characteristic curve (AUROC) at the minimum lambda is chosen. For scenario 1, the best alpha is 0.1. Two common lambdas are generally used: the lambda that gives the best performance (lambda.min) or the lambda with the fewest features selected that is within one standard error of the best-performing lambda (lambda.1se). We used lambda.1se, the most generalizable model, to avoid overfitting and selecting the most salient variables. From each of the cross-validations, a slightly different set of features were selected even though the features with the highest coefficients were relatively consistent. To account for the slight inconsistencies, the final set of features consists of features that were selected in all 10 cross-validations. The final set of features were used to predict estimated probabilities in the 10 test sets and calculate subsequent performance metrics.

Software

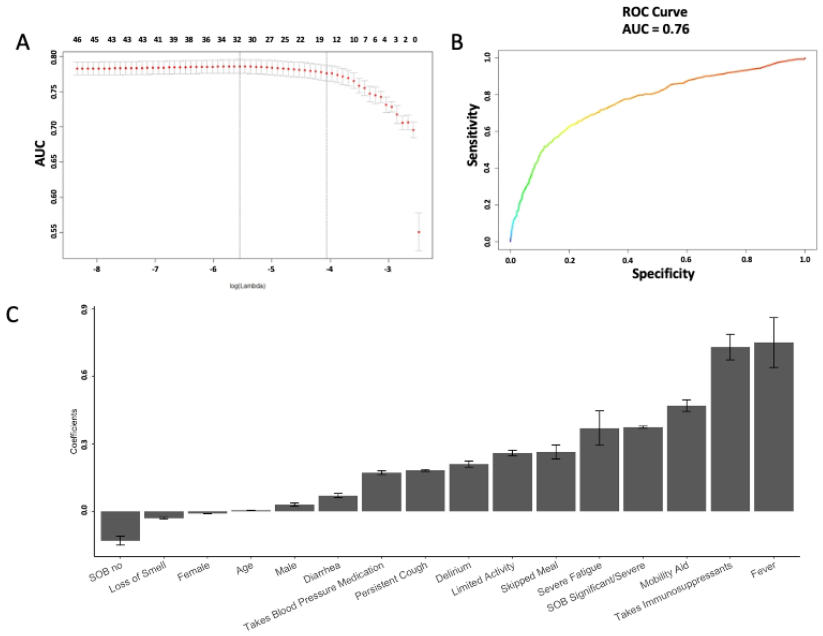
Analyses were all carried out in R (version 4.0.3). Packages used are listed as follows. For imputation, we used MICE (Multiple Imputation by Chained Equations; version 3.13.0). Imbalanced data set sampling was performed with ROSE (Random Over-Sampling Examples; version 0.0-3). For Elastic Net, we used glmnet (Lasso and Elastic-Net Regularized Generalized Linear Models; version 4.1).

Results

Elastic Net Results

We were able to predict users seeking treatment with relatively good accuracy. The average cross-validated area under the receiver operating characteristic curve (cvAUC) for the training set at the optimal parameters was 0.752 (SD 0.021). An example of the lambda path from one Elastic Net training is shown in Figure 1A, with the most parsimonious model selecting 16 features. Using the final set of features selected (Figure 1C) to calculate estimated probabilities and the area under the curve (AUC) on the 10 test sets, a similar accuracy was obtained, with an average AUC of 0.745 (SD 0.033). An example of AUC on the test set is shown in Figure 1B. The most important variables of this signature as selected by our predictive algorithm were fever, the use of immunosuppressant medication, the use of a mobility aid, shortness of breath, and severe fatigue. Age had a relatively small regression coefficient, indicating that pre-existing clinical conditions and symptom presentation are much stronger predictors of seeking treatment. Unexpectedly, BMI was not selected as a significant predictor. Finally, female gender was negatively associated with seeking treatment. Elastic Net regression was also applied to scenario 2. The prediction performance is comparable to scenario 1, and the selected features were also very similar (Figure S6 in Multimedia Appendix 1).

Figure 1. Elastic Net regression predictive performance and selected variables. We used Elastic Net regression where the outcome of seeking treatment or not was regressed on features in Tables 1-4. An example performance in terms of cross-validation AUC for Elastic Net classification on the training set across different values of lambda is shown in the top left. The final set of features selected from the Elastic Net model was applied on a holdout test data set and the resulting AUC is shown in the top right. The most important features selected by the Elastic Net model are shown at the bottom. Negative coefficients indicate a negative association with outcome and vice versa. AUC: area under the receiver operating characteristic curve; ROC: receiver operating characteristic; SOB: shortness of breath.



Logistic Regression Results

We next estimated the odds ratio from logistic regression using scenario 1 (Figure S4 in Multimedia Appendix 1). The most important features are consistent with the Elastic Net results. The modeling from logistic regression and Elastic Net regression using scenario 1 and 2 all selected similar features that are predictive of the outcome, lending robustness to the results.

Longitudinal Analysis

To better understand the fluctuations in the symptoms selected by the Elastic Net model, we then analyzed the symptoms in a longitudinal manner. We examined a window of time beginning 20 days before the patient goes to a hospital setting (for positive cases), and 20 days before the last entry (for negative cases); the results are shown in Figures 2-4. For each day, we estimated the frequency of each symptom for the positive and negative groups. Day 0 for the positive group corresponds to the day when the user seeks treatment, and day 0 for the negative group corresponds to the last entry. Figure 2 shows positive and negative groups of binary variables. Figure 3 shows categorical variables of fatigue and shortness of breath for the positive group, and Figure 4 shows fatigue and shortness of breath for the negative group. A linear regression line is superimposed for each group where the frequency is regressed on the days.

Figure 2. We analyzed all Elastic Net regression–selected symptoms for a 20-day window in a longitudinal manner, where day 0 for the positive group corresponds to the day when the user seeks treatment, and day 0 for the negative group corresponds to the last entry. The frequency of users having each feature for each day was plotted. A linear regression line where the frequency is regressed onto the days is plotted. The slopes and intercepts are labeled. Binary features of both positive and negative groups are plotted.

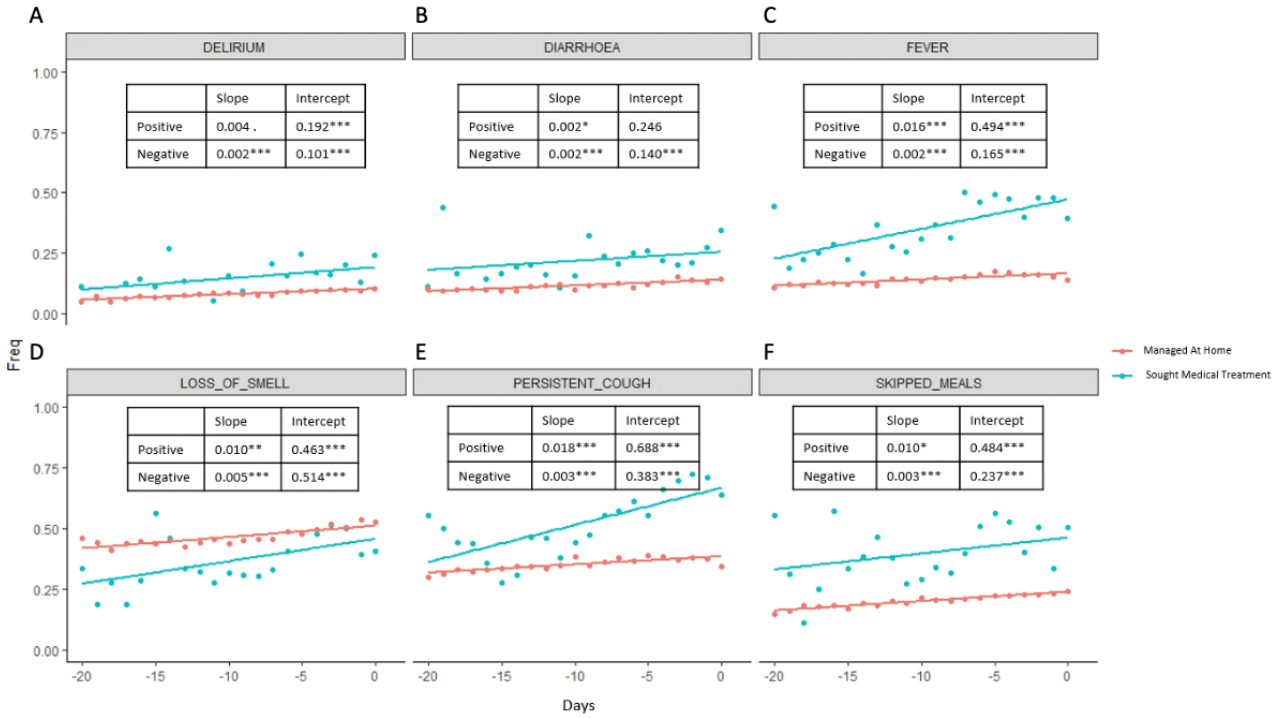


Figure 3. Categorical features of fatigue and shortness of breath are plotted for the positive group in longitudinal analysis.

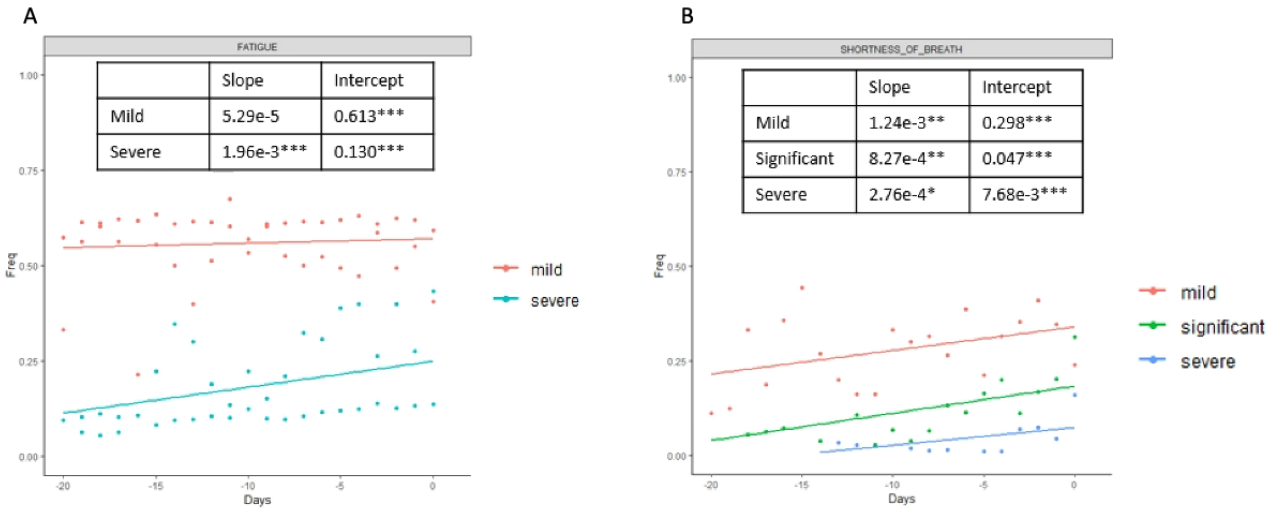
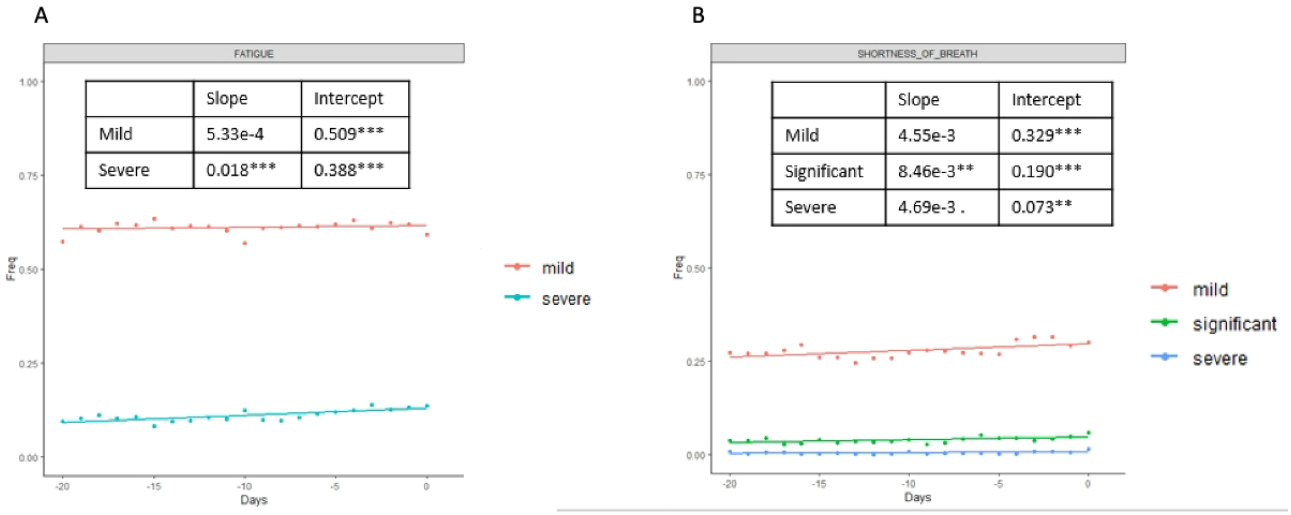


Figure 4. Categorical features of fatigue and shortness of breath are plotted for the negative group in longitudinal analysis.



Slope and intercepts are shown for comparison and their significance is evaluated using the likelihood ratio test (Figure S7 in Multimedia Appendix 1). The likelihood ratio test was used to compare whether there are statistically significant differences between the slopes of positive and negative cases in the trajectory analysis. Linear regression was used to quantify the association between days and frequency of each selected symptom in positive and negative cases. The likelihood ratio test was used to compare the linear regression model, where only the frequency of the feature was the independent variable, and the linear regression model, where the frequency of the feature and whether cases are positive or negative were the independent variables. The null hypothesis is that a linear model with only frequency of the feature as the independent variable is the superior model. The alternative hypothesis is that the superior model is the model with frequency of features and whether cases are positive or negative as independent variables. Rejection of the null hypothesis suggests that knowing the positive or negative cases predicts better frequency of the feature; therefore, the positive and negative cases are statistically different.

All differences between the two groups were significant except for mild fatigue. The slopes of the positive group were steeper than in the negative group in all the symptoms except for diarrhea, which indicates that the positive group had an increased frequency of symptoms that are indicative of severe COVID-19 cases as the disease progressed, while the frequency of the symptoms for the negative group stayed relatively stable. Not surprisingly, all the intercepts for the positive group are higher than for the negative group except for mild fatigue, further indicating that there are higher frequencies of COVID-19–related symptoms in users who were seeking treatment.

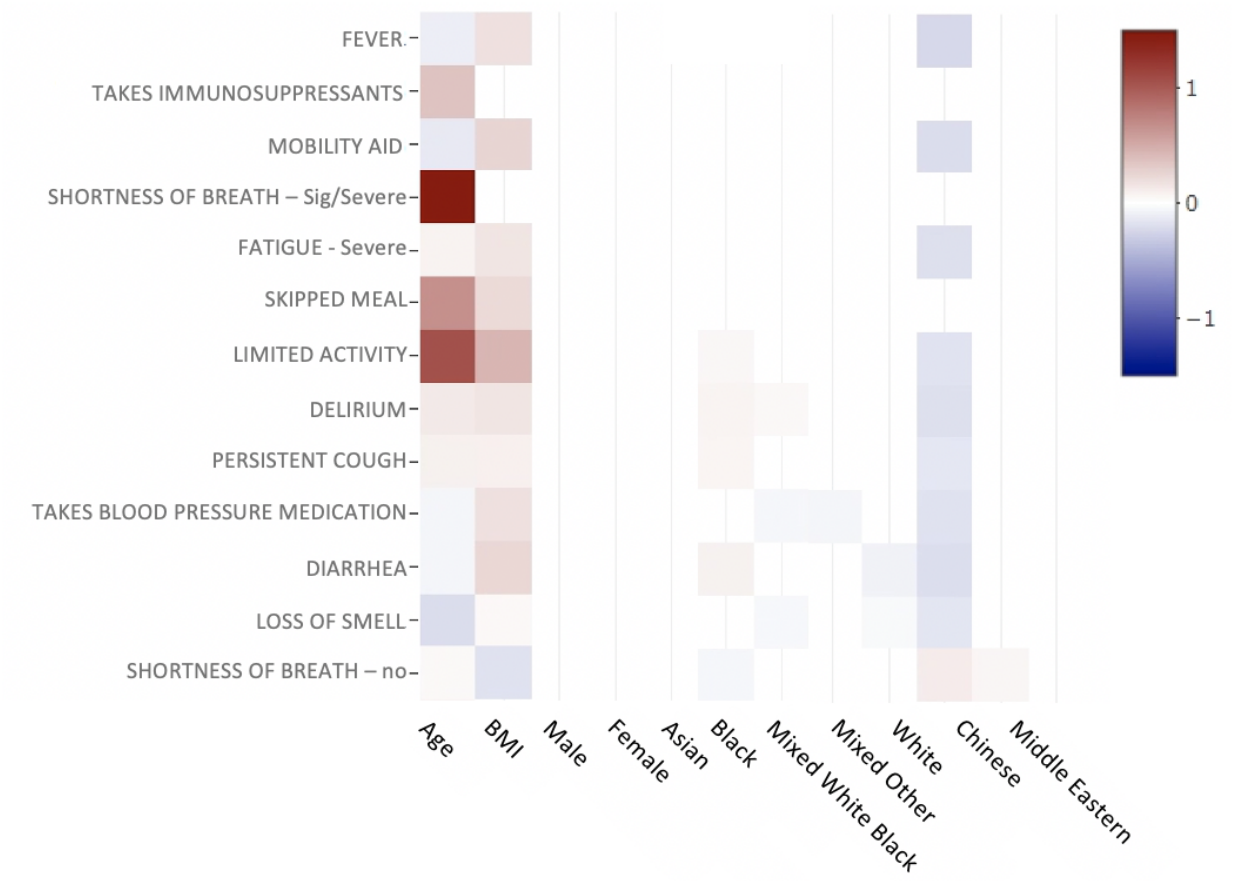
Age-Related Variables

To understand the age effects better, given that age has little significance in predicting the outcome, we analyzed the association between age and the other features selected. We conducted an experiment where we divided all the users who were positive for COVID-19 into three age groups: young, middle-aged, and older adults (aged <30 years, 30< aged <50 years, and aged >50 years, respectively). Running univariate logistic regression where the outcome of seeking treatment is regressed onto each feature selected by the Elastic Net model shows that the coefficients of the features do not vary substantially between age groups (Figure S5 in Multimedia Appendix 1). Such results suggest that the features' association with the outcome is not dependent on age.

SARS-CoV-2 has been shown to cause more severe disease in older adults [27]. Even though age was not a major contributor to the prediction of COVID-19–related treatment seeking, we explored whether age was associated with other features selected by the model. In conjunction, we also examined other demographic variables, such as race, BMI, and gender. We conducted multivariate logistic regression models where each of the features selected by the Elastic Net model was regressed on the demographic variables analyzed (Figure 5). Age was associated with 10/13 of the predictive features (P<.01). The most age-correlated features were use of a mobility aid, limited activity, blood pressure medication use, and immunosuppressant medication use. This indicates that age-related phenotypes in this cohort are associated with seeking treatment due to COVID-19. This emphasizes the fact that despite age, any population that expresses the features selected from our model could be susceptible to a more severe form of COVID-19. Understanding vulnerable young populations that are biologically older than their chronological age and exhibit

features that are generally associated with the older population could help identify susceptible young populations.

Figure 5. Multivariate logistic regression where each Elastic Net regression–selected feature is regressed onto demographic information such as age, BMI, gender, and race. Coefficients are plotted in a heat map. Only statistically significant associations are plotted. Age has a significant but weak association with many selected features. Users who identified as Black in the United Kingdom have many positive associations with high coefficients. Sig: significant.



In addition to age, being of Black ethnicity was associated with a number of features selected by the Elastic Net, such as a high frequency of delirium, limited activity, and blood pressure medication use. However, whether this is associated with socioeconomic status or an innate biological difference in people of African descent needs further investigation. The gender feature was a predictor of seeking treatment (Figure 1) but was not significantly correlated with any of the predictive features, suggesting that the sex of an individual affects other aspects of disease severity not evaluated in this study.

Discussion

In this paper, we used Elastic Net classifier to identify features that are pertinent to users seeking medical treatment for COVID-19. A relatively small effect of the loss of smell feature associated with mild disease outcomes was found (Figure 1). Such findings have also been reported in recent studies [28,29]. Female gender also had a negative correlation with seeking treatment, consistent with recent findings in large populations [30,31]. In our data, gender did not correlate strongly with any features, indicating that there may be factors other than comorbidities or symptoms that enable females to have a better prognosis. Underlying immunological differences in females [32-35] could lead to the mounting of a better immune response and neutralization of the virus more efficiently than in men.

From our analyses, we have found features that are predictive of people having a severe enough case of COVID-19 to be admitted into a hospital setting. However, there are some features where additional research would help elucidate the mechanism behind their correlation. For example, immunosuppressant use was a major predictor of a patient seeking treatment; however, from the data, we cannot investigate whether patients taking these drugs are more prone to severe COVID-19 because of underlying autoimmune or autoinflammatory disease or because of the direct effect of the drug on the suppression of the inflammatory response. If the latter was true, immunosuppressant use might be ameliorating severe disease phenotypes that are frequently caused by cytokine storms [36], which could be attenuated by the use of immunosuppressant medication.

Besides the need for additional research into the mechanism behind some of the features associated with a more severe disease state, time is an important variable that is not explored in depth in this paper. A Cox survival analysis would be informative; however, the start time of each user is inconsistent, making the application of Cox survival analysis inappropriate.

Some users' first entries already indicate testing positive for COVID-19, with symptoms suggesting that they are already in the middle of the disease course, while others slowly develop symptoms and test positive for COVID-19 later in time.

There are also limitations to a mobile app and using whether users seek treatment as a proxy to accurately capture "severity." Those who are able to manage their disease at home are considered less severe than those who sought treatment in a clinical or hospital setting. Such a proxy is different from the more formal severity scale developed by the World Health Organization. It is also conceivable that the sickest may not log their symptoms. However, we believe that self-reported data can provide more insight into when individuals with COVID-19 are sick enough to seek medical attention and what features are related to the change from self-care to medical treatment. Such higher resolution insight can provide those who have COVID-19 with warning signs of progressively worsening disease.

Age has been shown to be important in the severity of COVID-19 [13]. In our results, age shows a slight positive correlation with users seeking treatment. The difference between the average age of those who sought treatment and those who did not was relatively small, consistent with age not being a strong predictor. It is possible that the older population was less likely to use a smartphone app, leading to underrepresentation of the sick older population in our data set. To explore such a possibility, we accessed publicly available data on the age distribution of those infected with COVID-19 in the United Kingdom from March-September 2020 [37]. There is a bimodal distribution in the general UK population, with peaks at 50 and 80 years (average age of 57.6 years), while we only see a unimodal distribution in our data set, with the peak at 50 years (average age of 44.3 years; Figure S8 in Multimedia Appendix 1). A chi-square test comparing the COVID-19–positive population aged >60 years in the Tracker App versus the COVID-19–positive population aged >60 years in the UK general population also showed that there is a statistically significant difference (P<.001). This could explain the discrepancies between the small coefficient of importance for age in our model compared to previous literature, which were mostly published before September 2020, where age is a major factor in disease severity. Furthermore, the fact that age-associated variables outperform age in the prediction of a patient seeking treatment could also indicate that biological age or immunological age [38,39] may be appropriate measures for assessing an individual's prognosis. Chronological age is not a good representation of the condition of an individual. Biological age has been shown to be associated with all-cause mortality and susceptibility to diseases and may therefore be a more appropriate measure of how healthy an individual may be and how he/she might respond to COVID-19.

In conclusion, we identify age-dependent and age-independent sets of symptoms and comorbidities predictive of patients with COVID-19 seeking treatment. Our analyses show features that predict disease severity in advance, which could be used to predict severe cases of COVID-19 even in younger individuals who may not be labeled as high risk. A continued rise in the number of cases, as many societies struggle to balance reopening the economy and "flattening the curve," places an enormous burden on health care systems around the world. Knowing the signs of possible severe cases, including the ones derived in this study, could help health care systems devote resources to intervening in potentially severe cases before they become costly to manage.

Acknowledgments

This work uses data provided by participants of the COVID-19 Symptoms Study, developed by ZOE Global Limited with scientific and clinical input from King's College London. We would also like to acknowledge all data providers who made anonymized data available for research.

We wish to acknowledge the collaborative partnership that enabled acquisition of and access to the deidentified data that led to this output. The collaboration was led by BREATHE - The Health Data Research Hub for Respiratory Health, in partnership with Secure Anonymised Information Linkage (SAIL) Databank. We wish to acknowledge the input of ZOE Global Limited and King's College London in their development and sharing of the data, and their input into the understanding and contextualization of the data for COVID-19 research. All research conducted has been completed under the permission and approval of SAIL's independent Information Governance Review Panel (project number 1088, project lead Dr Dina Radenkovic).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 4817 KB - jmir_v23i3e25696_app1.docx]

References

1. Halacli B, Kaya A, Topeli A. Critically-ill COVID-19 patient. Turk J Med Sci 2020 Apr 21;50(SI-1):585-591 [FREE Full text] [doi: 10.3906/sag-2004-122] [Medline: 32293831]

2. Lin Y. [Intensive Care During a Global Epidemic]. Hu Li Za Zhi 2020 Jun;67(3):4-5. [doi: 10.6224/JN.202006_67(3).01] [Medline: 32495323]
3. Armstrong RA, Kane AD, Cook TM. Outcomes from intensive care in patients with COVID-19: a systematic review and meta-analysis of observational studies. Anaesthesia 2020 Oct 15;75(10):1340-1349 [FREE Full text] [doi: 10.1111/anae.15201] [Medline: 32602561]
4. Bartsch SM, Ferguson MC, McKinnell JA, O'Shea KJ, Wedlock PT, Siegmund SS, et al. The Potential Health Care Costs And Resource Use Associated With COVID-19 In The United States. Health Aff (Millwood) 2020 Jun 01;39(6):927-935. [doi: 10.1377/hlthaff.2020.00426] [Medline: 32324428]
5. Noronha KDS, Guedes G, Turra C, Andrade M, Botega L, Nogueira D, et al. The COVID-19 pandemic in Brazil: analysis of supply and demand of hospital and ICU beds and mechanical ventilators under different scenarios. Cad Saude Publica 2020;36(6):e00115320 [FREE Full text] [doi: 10.1590/0102-311X00115320] [Medline: 32578805]
6. Weissman GE, Crane-Droesch A, Chivers C, Luong T, Hanish A, Levy MZ, et al. Locally Informed Simulation to Predict Hospital Capacity Needs During the COVID-19 Pandemic. Annals of Internal Medicine 2020 Jul 07;173(1):21-28. [doi: 10.7326/m20-1260]
7. Khera R, Jain S, Lin Z, Ross J, Krumholz H. Evaluation of the Anticipated Burden of COVID-19 on Hospital-Based Healthcare Services Across the United States. medRxiv. Preprint published online on April 3, 2020. [FREE Full text] [doi: 10.1101/2020.04.01.20050492] [Medline: 32511584]
8. Carenzo L, Costantini E, Greco M, Barra FL, Rendiniello V, Mainetti M, et al. Hospital surge capacity in a tertiary emergency referral centre during the COVID-19 outbreak in Italy. Anaesthesia 2020 Jul 22;75(7):928-934. [doi: 10.1111/anae.15072] [Medline: 32246838]
9. Moghadas SM, Shoukat A, Fitzpatrick MC, Wells CR, Sah P, Pandey A, et al. Projecting hospital utilization during the COVID-19 outbreaks in the United States. Proc Natl Acad Sci U S A 2020 Apr 21;117(16):9122-9126 [FREE Full text] [doi: 10.1073/pnas.2004064117] [Medline: 32245814]
10. Sun Q, Qiu H, Huang M, Yang Y. Lower mortality of COVID-19 by early recognition and intervention: experience from Jiangsu Province. Ann Intensive Care 2020 Mar 18;10(1):33 [FREE Full text] [doi: 10.1186/s13613-020-00650-2] [Medline: 32189136]
11. Chen M, Tu C, Tan C, Zheng X, Wang X, Wu J, et al. Key to successful treatment of COVID-19: accurate identification of severe risks and early intervention of disease progression. medRxiv. Preprint published online on April 11, 2020. [doi: 10.1101/2020.04.06.20054890]
12. Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, COPE Consortium. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. Science 2020 Jun 19;368(6497):1362-1367 [FREE Full text] [doi: 10.1126/science.abc0473] [Medline: 32371477]
13. Garg S, Kim L, Whitaker M, O'Halloran A, Cummings C, Holstein R, et al. Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14 States, March 1-30, 2020. MMWR Morb Mortal Wkly Rep 2020 Apr 17;69(15):458-464 [FREE Full text] [doi: 10.15585/mmwr.mm6915e3] [Medline: 32298251]
14. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, Centre for Mathematical Modelling of Infectious Diseases COVID-19 working group. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 2020 May;20(5):553-558 [FREE Full text] [doi: 10.1016/S1473-3099(20)30144-4] [Medline: 32171059]
15. Kang SJ, Jung SI. Age-Related Morbidity and Mortality among Patients with COVID-19. Infect Chemother 2020 Jun;52(2):154-164 [FREE Full text] [doi: 10.3947/ic.2020.52.2.154] [Medline: 32537961]
16. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
17. Dhochak N, Singhal T, Kabra SK, Lodha R. Pathophysiology of COVID-19: Why Children Fare Better than Adults? Indian J Pediatr 2020 Jul 14;87(7):537-546 [FREE Full text] [doi: 10.1007/s12098-020-03322-y] [Medline: 32410003]
18. Wang W, Tang J, Wei F. Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. J Med Virol 2020 Apr 12;92(4):441-447 [FREE Full text] [doi: 10.1002/jmv.25689] [Medline: 31994742]
19. Jones KH, Ford DV, Jones C, Dsilva R, Thompson S, Brooks CJ, et al. A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: a privacy-protecting remote access system for health-related research and evaluation. J Biomed Inform 2014 Aug;50:196-204 [FREE Full text] [doi: 10.1016/j.jbi.2014.01.003] [Medline: 24440148]
20. Ford DV, Jones KH, Verplancke J, Lyons RA, John G, Brown G, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. BMC Health Serv Res 2009 Sep 04;9(1):157 [FREE Full text] [doi: 10.1186/1472-6963-9-157] [Medline: 19732426]
21. Lyons RA, Jones KH, John G, Brooks CJ, Verplancke J, Ford DV, et al. The SAIL databank: linking multiple health and social care datasets. BMC Med Inform Decis Mak 2009 Jan 16;9(1):3 [FREE Full text] [doi: 10.1186/1472-6947-9-3] [Medline: 19149883]

22. Rodgers SE, Demmler JC, Dsilva R, Lyons RA. Protecting health data privacy while using residence-based environment and demographic data. Health Place 2012 Mar;18(2):209-217. [doi: 10.1016/j.healthplace.2011.09.006] [Medline: 21996431]
23. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J Stat Soft 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
24. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, et al. Missing data and multiple imputation in clinical epidemiological research. CLEP 2017 Mar;Volume 9:157-166. [doi: 10.2147/clep.s129785]
25. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009 Jun 29;338(jun29 1):b2393-b2393 [FREE Full text] [doi: 10.1136/bmj.b2393] [Medline: 19564179]
26. Bellinger C, Sharma S, Zaı¨ane OR, Japkowicz N. Sampling a Longer Life: Binary versus One-class classification Revisited. 2017 Presented at: First International Workshop on Learning with Imbalanced Domains: Theory and Applications; September 22, 2017; Skopje, Macedonia p. 64-78 URL: http://proceedings.mlr.press/v74/bellinger17a/bellinger17a.pdf
27. Shahid Z, Kalayanamitra R, McClafferty B, Kepko D, Ramgobin D, Patel R, et al. COVID-19 and Older Adults: What We Know. J Am Geriatr Soc 2020 May 20;68(5):926-929 [FREE Full text] [doi: 10.1111/jgs.16472] [Medline: 32255507]
28. Yan CH, Faraji F, Prajapati DP, Ostrander BT, DeConde AS. Self-reported olfactory loss associates with outpatient clinical course in COVID-19. Int Forum Allergy Rhinol 2020 Jul 07;10(7):821-831 [FREE Full text] [doi: 10.1002/alr.22592] [Medline: 32329222]
29. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul 11;26(7):1037-1040 [FREE Full text] [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
30. Jin J, Bai P, He W, Wu F, Liu XF, Han DM, et al. Gender Differences in Patients With COVID-19: Focus on Severity and Mortality. Front Public Health 2020;8:152 [FREE Full text] [doi: 10.3389/fpubh.2020.00152] [Medline: 32411652]
31. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, the Northwell COVID-19 Research Consortium, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA 2020 May 26;323(20):2052-2059 [FREE Full text] [doi: 10.1001/jama.2020.6775] [Medline: 32320003]
32. Park MD. Sex differences in immune responses in COVID-19. Nat Rev Immunol 2020 Aug 22;20(8):461-461 [FREE Full text] [doi: 10.1038/s41577-020-0378-2] [Medline: 32572247]
33. Scully EP, Haverfield J, Ursin RL, Tannenbaum C, Klein SL. Considering how biological sex impacts immune responses and COVID-19 outcomes. Nat Rev Immunol 2020 Jul 11;20(7):442-447 [FREE Full text] [doi: 10.1038/s41577-020-0348-8] [Medline: 32528136]
34. Takahashi T, Ellingson MK, Wong P, Israelow B, Lucas C, Klein J, Yale IMPACT Research Team, et al. Sex differences in immune responses that underlie COVID-19 disease outcomes. Nature 2020 Dec;588(7837):315-320 [FREE Full text] [doi: 10.1038/s41586-020-2700-3] [Medline: 32846427]
35. Furman D. Sexual dimorphism in immunity: improving our understanding of vaccine immune responses in men. Expert Rev Vaccines 2015 Mar 02;14(3):461-471. [doi: 10.1586/14760584.2015.966694] [Medline: 25278153]
36. Ye Q, Wang B, Mao J. The pathogenesis and treatment of the 'Cytokine Storm' in COVID-19. J Infect 2020 Jun;80(6):607-613 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.037] [Medline: 32283152]
37. Official UK Coronavirus Dashboard. URL: https://coronavirus.data.gov.uk/ [accessed 2021-01-15]
38. Alpert A, Pickman Y, Leipold M, Rosenberg-Hasson Y, Ji X, Gaujoux R, et al. A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. Nat Med 2019 Mar 6;25(3):487-495 [FREE Full text] [doi: 10.1038/s41591-019-0381-y] [Medline: 30842675]
39. Sayed N, Gao T, Tibshirani R, Hastie T, Cui L, Kuznetsova T, et al. An Inflammatory Clock Predicts Multi-morbidity, Immunosenescence and Cardiovascular Aging in Humans. bioRxiv. Preprint published online on November 16, 2019. [doi: 10.1101/840363]

Abbreviations

AUC: area under the curve
AUROC: area under the receiver operating characteristic curve
cvAUC: cross-validated area under the receiver operating characteristic curve
ICU: intensive care unit
WHO: World Health Organization

Edited by G Eysenbach; submitted 14.11.20; peer-reviewed by P Giabbanelli, Y Chu; comments to author 01.01.21; revised version received 20.01.21; accepted 08.02.21; published 25.03.21.

Please cite as:
Huang Y, Radenkovic D, Perez K, Nadeau K, Verdin E, Furman D
Modeling Predictive Age-Dependent and Age-Independent Symptoms and Comorbidities of Patients Seeking Treatment for COVID-19: Model Development and Validation Study
J Med Internet Res 2021;23(3):e25696
URL: https://www.jmir.org/2021/3/e25696
doi:10.2196/25696
PMID:33621185

Original Paper

Understanding Concerns, Sentiments, and Disparities Among Population Groups During the COVID-19 Pandemic Via Twitter Data Mining: Large-scale Cross-sectional Study

Chunyan Zhang[1*], MS; Songhua Xu[1*], PhD; Zongfang Li[1*], PhD; Shunxu Hu[2], BS

[1]Institute of Medical Artificial Intelligence, The Second Affiliate Hospital of Xi'an Jiaotong University, Xi'an, China

[2]School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China

[*]these authors contributed equally

Corresponding Author:
Songhua Xu, PhD
Institute of Medical Artificial Intelligence
The Second Affiliate Hospital of Xi'an Jiaotong University
No.157 Xiwu Road
Xi'an
China
Phone: 86 18710823698
Email: songhua_xu1@163.com

Abstract

Background: Since the beginning of the COVID-19 pandemic in late 2019, its far-reaching impacts have been witnessed globally across all aspects of human life, such as health, economy, politics, and education. Such widely penetrating impacts cast significant and profound burdens on all population groups, incurring varied concerns and sentiments among them.

Objective: This study aims to identify the concerns, sentiments, and disparities of various population groups during the COVID-19 pandemic through a cross-sectional study conducted via large-scale Twitter data mining infoveillance.

Methods: This study consisted of three steps: first, tweets posted during the pandemic were collected and preprocessed on a large scale; second, the key population attributes, concerns, sentiments, and emotions were extracted via a collection of natural language processing procedures; third, multiple analyses were conducted to reveal concerns, sentiments, and disparities among population groups during the pandemic. Overall, this study implemented a quick, effective, and economical approach for analyzing population-level disparities during a public health event. The source code developed in this study was released for free public use at GitHub.

Results: A total of 1,015,655 original English tweets posted from August 7 to 12, 2020, were acquired and analyzed to obtain the following results. Organizations were significantly more concerned about COVID-19 (odds ratio [OR] 3.48, 95% CI 3.39-3.58) and expressed more fear and depression emotions than individuals. Females were less concerned about COVID-19 (OR 0.73, 95% CI 0.71-0.75) and expressed less fear and depression emotions than males. Among all age groups (ie, ≤18, 19-29, 30-39, and ≥40 years of age), the attention ORs of COVID-19 fear and depression increased significantly with age. It is worth noting that not all females paid less attention to COVID-19 than males. In the age group of 40 years or older, females were more concerned than males, especially regarding the economic and education topics. In addition, males 40 years or older and 18 years or younger were the least positive. Lastly, in all sentiment analyses, the sentiment polarities regarding political topics were always the lowest among the five topics of concern across all population groups.

Conclusions: Through large-scale Twitter data mining, this study revealed that meaningful differences regarding concerns and sentiments about COVID-19-related topics existed among population groups during the study period. Therefore, specialized and varied attention and support are needed for different population groups. In addition, the efficient analysis method implemented by our publicly released code can be utilized to dynamically track the evolution of each population group during the pandemic or any other major event for better informed public health research and interventions.

(J Med Internet Res 2021;23(3):e26482) doi:10.2196/26482

KEYWORDS

COVID-19; Twitter mining; infodemiology; infoveillance; pandemic; concerns; sentiments; population groups; disparities

Introduction

Background

Since December 2019, COVID-19 has rapidly spread all over the world and caused millions of deaths [1,2]. Although many countries have implemented various countermeasures [3,4], an end to the pandemic is still not in sight. So far, COVID-19 has already exerted tremendous impacts across various aspects of human life, such as health, economy, politics, and education [5-8], whose influences may last for an unknown period. Such widely penetrating and long-lasting impacts are likely to cause disproportionate burdens on different population groups, incurring varied concerns and sentiments among them. Therefore, it is of great importance to understand the disparities in the responses of these population groups to COVID-19 for better informed public health research and intervention.

Literature Reviews

So far, two classes of methods have been utilized to study the impacts of COVID-19 on public and personal life, including large-scale social media mining approaches and cross-sectional analyses through online and offline questionnaires, which are briefly reviewed in the following text.

The first class of methods provides a fast and economical way to analyze the population impacts of COVID-19 through mining social media data generated during the pandemic. Currently, such methods have been employed in a number of studies. For example, Lwin et al [9] studied Twitter data to explore global trends of four emotions—fear, anger, sadness, and joy—as well as their relative salience. After studying the topics obtained by latent Dirichlet allocation topic modeling on Twitter text data, Abd-Alrazaq et al [10] identified the sentiments of four major topics and 12 subtopics, and showed that all topics were positive except for two (ie, death and racial discrimination). Similarly, Hung et al [11] adopted the Valence Aware Dictionary and Emotional Reasoner (VADER) model to analyze the sentiments expressed in user tweets and found that positive, neutral, and negative emotions accounted for 48.2%, 20.7%, and 31.1% of the tweets, respectively.

Despite the informative understanding regarding people's sentiments provided by these prior studies, it is noted that these existing methods tend to treat their study population as a whole in the analysis, ignoring likely disparities among population groups. Case reports from many countries and epidemiological research on COVID-19 state that the morbidity and mortality of COVID-19 are related to age and gender [12-14], calling for a more fine-grained analysis regarding the concerns and sentiments of each population group during the pandemic.

The second class of methods has been popularly leveraged to understand the health statuses of population groups, uncover health-related factors, and carry out disease epidemiology research. Table 1 [15-20] lists some representative cross-sectional surveys on COVID-19. Compared with the first class of data mining methods, cross-sectional studies can provide richer and more fine-grained information through well-controlled questionnaires, which is of great use for analyzing the detailed disparities of population groups.

Table 1. Representative cross-sectional studies on COVID-19.

Author and reference	Study target area	Study period (all in 2020)	No. of participants (online or offline)	Highlights
Liu et al [15]	Wuhan and surrounding cities, China	January 30-February 8	300 (online)	Gender differences exist in posttraumatic stress symptoms during COVID-19: females suffer more than males.
Lu et al [16]	Fujian, China	April 6-22	2299 (offline)	Work differences exist in fear, anxiety, and depression emotions in hospitals during COVID-19: medical workers suffer more than administrative workers.
Nelson et al [17]	Parts of the United States	March 14-16	9009 (online)	Age differences exist in concerns about COVID-19: people aged 40-54 years and 55-75 years are very worried and extremely worried population groups, respectively.
Groarke et al [18]	The United Kingdom	March 23-April 24	1964 (online)	Age differences exist in loneliness during COVID-19: young people suffer most.
Azlan et al [19]	Malaysia	March 27-April 3	4850 (online)	Gender, age, region, occupation, and income differences exist in public knowledge toward COVID-19.
Ahmad and Murad [20]	Iraqi Kurdistan	Not stated	516 (online)	Age differences exist in mental health during COVID-19: young people aged 18-35 years are facing psychological anxiety.

However, the shortcomings of both online and offline cross-sectional studies are also commonly acknowledged. In particular, launching offline questionnaires during the COVID-19 pandemic may pose eminent public health hazards because of the risk of virus transmission through personal contacts. Online questionnaires also have their own challenges, mainly difficulties in finding an adequate number of willing participants to complete the online questionnaires honestly and

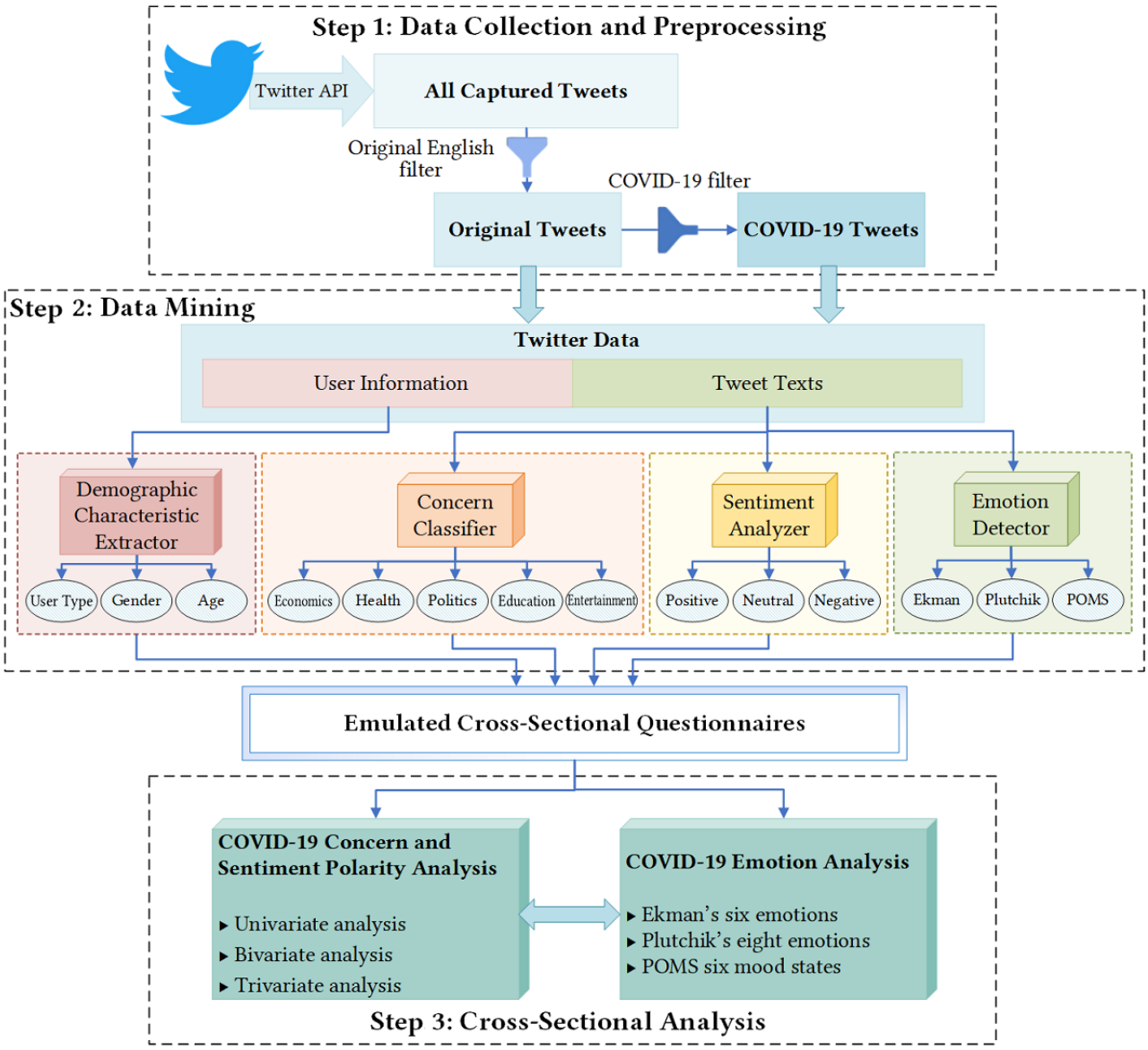
at a high quality. The operational obstacle of online questionnaires is further elevated if repeated surveys are intended to track the dynamic evolution of population groups regarding their thoughts and needs [21].

Recognizing the limitations of the two classes of existing study methods, in this work, we conducted a new cross-sectional study via large-scale Twitter data mining. Through this method, we aimed to identify the concerns, sentiments, and disparities of various population groups during the COVID-19 pandemic in fine granularity without administrating any online or offline questionnaires. The advantage of our approach lies in its economic and efficient way of gathering multifaceted awareness information from population groups and their disparities. With such an understanding of the concerns and sentiments of population groups regarding COVID-19, specialized attention and customized programs can be developed to assist each population group. It is noted that the method implemented through our social media data mining approach can be easily repurposed to study the evolution of different population groups during any major public health event for better informed public health research and interventions. The source code developed in this study has been released for free public use at GitHub [22].

Methods

As shown in Figure 1, the cross-sectional method proposed in this study consists of three steps. The implementation details of each step are described in the following sections.

Figure 1. The structure of our cross-sectional method. API: application programming interface; POMS: Profile of Mood States.



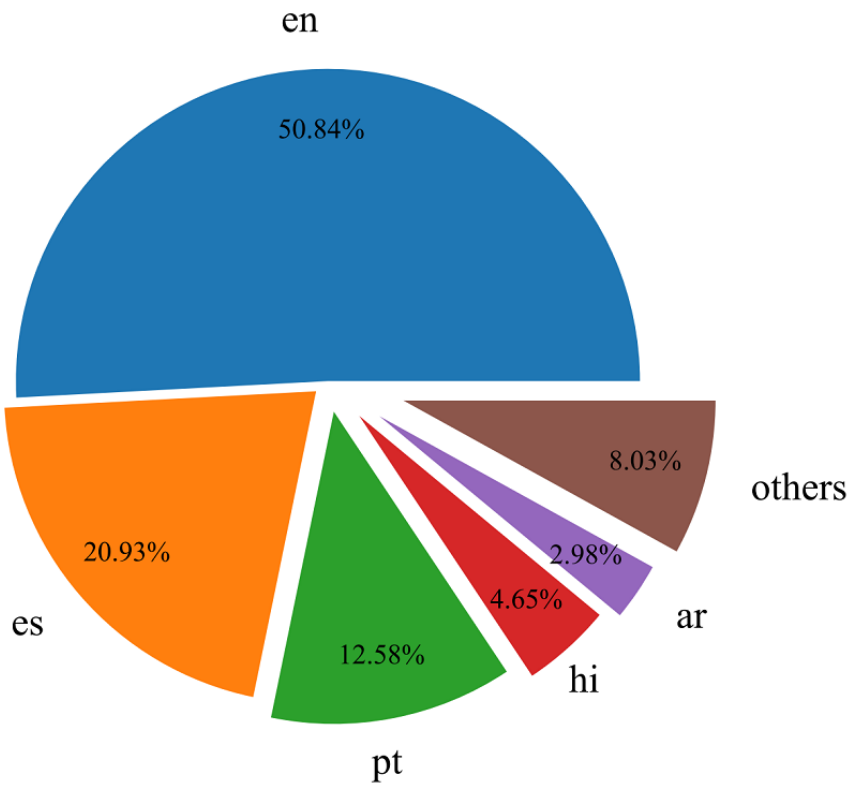
Data Collection and Preprocessing

The Twitter data used in this study were collected by sampled stream application programming interface v1 [23] and v2 [24] from Twitter Developer Labs, which can stream about 1% of publicly available tweets in real time. Meanwhile, detailed author data from all the tweets were collected to extract population characteristics. Unlike those in other research studies on Twitter [9-11], the data captured in this study are a random sampling of all Twitter data without using any filter, which can

better reflect the common opinions in people's daily lives. As of November 2020, we have collected, in total, more than 600 million tweets (ie, over 2 Terabytes) during the COVID-19 pandemic.

In the data preprocessing step, an original English filter and a COVID-19 filter were used to generate the original and COVID-19 tweet data sets based on all the captured tweets. Since original tweets can better reflect the authors' dynamic thoughts and sentiments, and English tweets comprise over half of all tweets (see Figure 2), we only focused on original English tweets, which can be filtered by the attributes of the tweet object. In order to obtain COVID-19 tweets, we made a filter pattern that is composed of 590 COVID-19 keywords and hashtags provided by Twitter COVID-19 filter rules [25].

Figure 2. The language distribution of tweets. ar: Arabic; en: English; es: Spanish; hi: Hindi; others: other languages; pt: Portuguese.



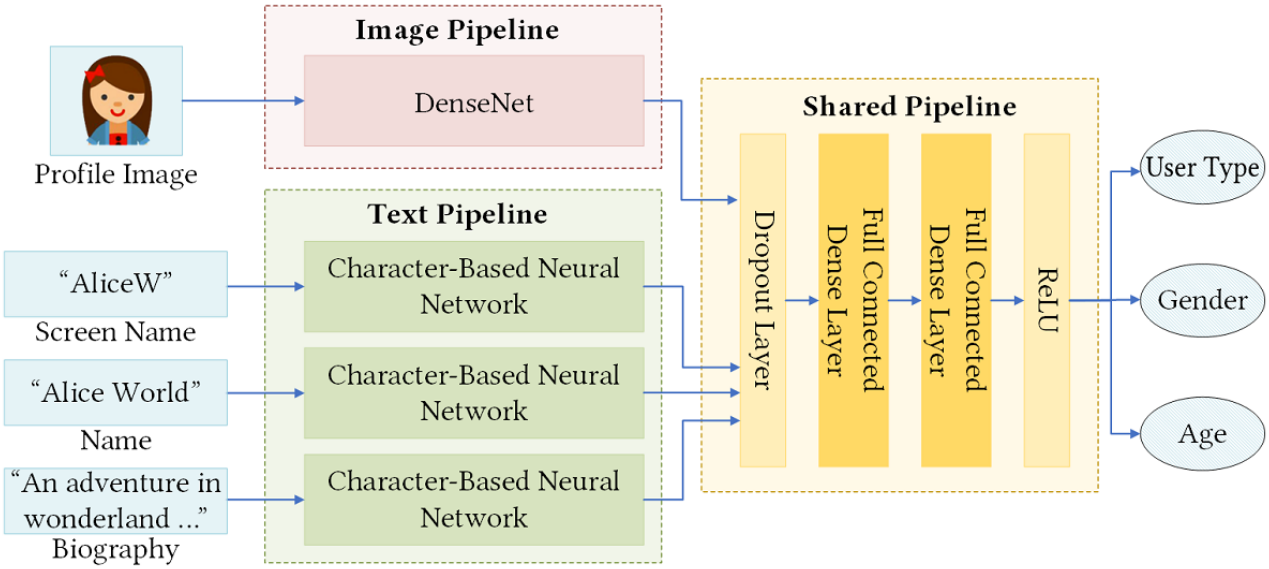
Data Mining

Data mining is the key step in emulating cross-sectional questionnaires based on the two tweet data sets. This step contained four intelligent modules: demographic characteristic extractor, concern classifier, sentiment analyzer, and emotion detector.

Demographic Characteristic Extractor

This module was used to extract three demographic characteristics—user type, gender, and age—through profile images, screen names, names, and biographies. It was implemented by an open source package of the M3 (multimodal, multilingual, and multi-attribute) model [26], which is a multimodal deep neural system trained on a massive data set, composed of Twitter, IMDB, and Wikipedia data [27], for demographic inference. In this M3 model, user type (ie, person or organization) and gender (ie, male or female) were modeled as binary classification tasks, while age was modeled as a 4-class classification task with the following age groups: ≤18, 19-29, 30-39, and ≥40 years of age. As shown in Figure 3, the structure of the M3 model consisted of two separate pipelines—image pipeline and text pipeline—and a shared pipeline. The image pipeline was employed to process profile images using the dense convolutional network (DenseNet) [28], and the text pipeline was used for processing three text sources of screen names, names, and biographies by adopting three character-based neural networks. The shared pipeline combined the outputs of the two separate pipelines and then mainly applied two fully connected dense layers to predict the user type, gender, and age state of each Twitter user. All of these pipelines were fine-tuned to capture accurate demographic features. For more detailed information, readers can refer to the original literature [26].

Figure 3. The structure of the M3 (multimodal, multilingual, and multi-attribute) model for inferring user type, gender, and age from profile information. DenseNet: dense convolutional network; ReLU: rectified linear unit.



We tested the M3 model on a subset of our original English tweets that carried ground-truth labels of user type, gender, and age explicitly or implicitly; the detection procedure is explained in detail in Multimedia Appendix 1. The benchmark performance of the M3 model on this subset is as follows: for user type, gender, and age, the accuracy scores are 99.07%, 95.88%, and 77.65%, respectively, and the macro–F1 scores are 0.9860, 0.9572, and 0.7311, respectively.

Concern Classifier

This module was used to classify the tweets into five categories of human life—economics, politics, health, education, and entertainment—which was based on our self-designed matching patterns. First, five specialized vocabulary dictionaries were collected and constructed from Oxford Reference and other sources, including an economic vocabulary (ie, A Dictionary of Economics [29] and The Economist [30]) and a political vocabulary (ie, A Concise Oxford Dictionary of Politics and International Relations [31]). Then, the vocabulary dictionaries were imported into the matching patterns in a regular expression format, with which we labeled all the tweets.

Sentiment Analyzer

This module calculated the sentiment polarities of the tweets based on the VADER [32] model. The VADER model is a sentiment analysis tool based on lexicons of sentiment-related words, which can automatically classify each word in the lexicon as positive, neutral, or negative. The range of the sentiment polarity is –1 to 1, which is divided into three subranges: negative (–1 to –0.05), neutral (–0.05 to 0.05), and positive (0.05 to 1).

Emotion Detector

This module is based on an emotion recognition model on Twitter [33], which utilizes a character-based trained recurrent neural network algorithm. It employs three emotion models to recognize different human emotions, including Ekman's six basic emotions model [34]; Plutchik's eight primary emotions model, also known as the emotion wheel [35]; and the Profile of Mood States (POMS) model [36], which measures six mood states. Based on the above-mentioned modules, the template of the emulated cross-sectional questionnaire is shown in Table 2.

Table 2. The template of the cross-sectional questionnaire.

Question category	Response categories
Population characteristic	
User type	Person
	Organization
Gender	Male
	Female
Age (years)	≤18
	19-29
	30-39
	≥40
Concern	
Economics	Concerned
	Unconcerned
Health	Concerned
	Unconcerned
Politics	Concerned
	Unconcerned
Education	Concerned
	Unconcerned
Entertainment	Concerned
	Unconcerned
Sentiment polarity	
Negative	–1 to –0.05
Neutral	–0.05 to 0.05
Positive	0.05 to 1
Emotions	
Ekman's six emotions: anger, disgust, fear, joy, sadness, and surprise	0 to 1 for each emotion
Plutchik's eight emotions: anger, disgust, fear, joy, sadness, surprise, trust, and anticipation	0 to 1 for each emotion
POMS[a] six emotions: anger, depression, fatigue, vigor, tension, and confusion	0 to 1 for each emotion

[a]POMS: Profile of Mood States.

Cross-sectional Analysis

The purpose of this step was to analyze the concerns and sentiments of different population groups in response to COVID-19 based on the Twitter data mining outcomes of the emulated questionnaire. It includes two parts: the COVID-19 concern and sentiment polarity analysis and the COVID-19 emotion analysis. The odds ratio (OR) was employed in these two parts to compare the relative ratios of population groups under multiple variable conditions. Meanwhile, we used the chi-square test to measure the significance level of difference (ie, *P* value) under each condition.

Results

Overall Analysis

During the COVID-19 pandemic, various emotions were expressed by the general public. To study the disparities between different population groups during this period, we conducted a cross-sectional analysis on the daily Twitter data collected from August 7 to 12, 2020. In total, 7,590,844 unfiltered tweets were captured during the research period, of which 1,015,655 were original English tweets; these are referred to as the original data set. From this original data set, 27,216 tweets were related to COVID-19; these are referred to as COVID-19 data set. The statistical distributions and *P* values, by chi-square test, of the two data sets are shown in Table 3.

We can see from Table 3 that the population groups under each variable all showed significant differences ($P<.001$) in response

to COVID-19. As shown in Table 3, 89.94% of the total participants were persons and 10.06% were organizations. As a comparison, 73.00% and 27.00% of COVID-19-related participants were persons and organizations, respectively. The total proportion of male participants on social media was slightly higher than that of females (52.74% vs 47.26%), while this gap was further widened to 60.38% versus 39.62% under COVID-19, respectively. The total proportions of the four age groups—≤18, 19-29, 30-39, and ≥40 years of age—were 37.93%, 38.42%, 11.41%, and 12.24%, respectively; from this, it can be inferred that people below 30 years of age are more active on social media. Under COVID-19, the proportions increased in the age groups above 30 years and decreased in the age groups below 30 years; thus, the proportions of the four age groups changed to 17.83%, 29.18%, 18.32%, and 34.67%, respectively. The total proportions of the five topics—economics, health, politics, education, and entertainment—were 13.99%, 13.90%, 7.27%, 6.38%, and 7.79%, respectively; under COVID-19, their proportions changed to 34.30%, 22.60%, 19.97%, 15.74%, and 6.38%, respectively. The total proportions of positive, neutral, and negative sentiments were 42.46%, 31.38%, and 26.16%, respectively; the mean sentiment polarity was 0.1067 (SD 0.4647). Under COVID-19, the proportions of positive, neutral, and negative sentiments were 43.15%, 24.37%, and 32.48%, respectively; the mean sentiment polarity fell to 0.0659 (SD 0.4941).

Table 3. Statistical distributions of the emulated questionnaire answers.

Variable	Total tweets, n (%)[a]	COVID-19-related tweets, n (%)	*P* value
Overall	1,015,655 (100)	27,216 (100)	N/A[b]
User type			
Person	913,480 (89.94)	19,869 (73.00)	<.001
Organization	102,175 (10.06)	7347 (27.00)	N/A
Gender			
Male	481,770 (52.74)	11,997 (60.38)	<.001
Female	431,710 (47.26)	7872 (39.62)	N/A
Age (years)			
≤18	346,483 (37.93)	3542 (17.83)	<.001
19-29	350,959 (38.42)	5798 (29.18)	N/A
30-39	104,228 (11.41)	3640 (18.32)	N/A
≥40	111,810 (12.24)	6889 (34.67)	N/A
Concern			
Economics	142,090 (13.99)	9334 (34.30)	<.001
Health	141,176 (13.90)	6152 (22.60)	N/A
Politics	73,838 (7.27)	5434 (19.97)	N/A
Education	64,799 (6.38)	4284 (15.74)	N/A
Entertainment	79,119 (7.79)	1736 (6.38)	N/A
Sentiment polarity			
Overall (–1 to 1), mean (SD)	0.1067 (0.4647)	0.0659 (0.4941)	<.001
Positive (–1 to 0.05)	431,247 (42.46)	11,744 (43.15)	<.001
Neutral (–0.05 to 0.05)	318,713 (31.38)	6632 (24.37)	N/A
Negative (0.05 to 1)	265,695 (26.16)	8840 (32.48)	N/A

[a]All values are expressed as n (%), except for overall sentiment polarity, which is expressed as mean (SD).

[b]*P* values were calculated for the main variables and not for individual responses.

The above analysis cannot provide fine-grained differences between population groups under multivariate conditions. To understand these differences more clearly, we adopted a cross-sectional analysis based on the emulated questionnaire outcomes, which consists of two parts: one is COVID-19 concern and sentiment polarity analysis, including univariate, bivariate, and trivariate analysis, and the other one is COVID-19 emotion analysis, including three emotion models. The analysis process and results are presented in the following sections.

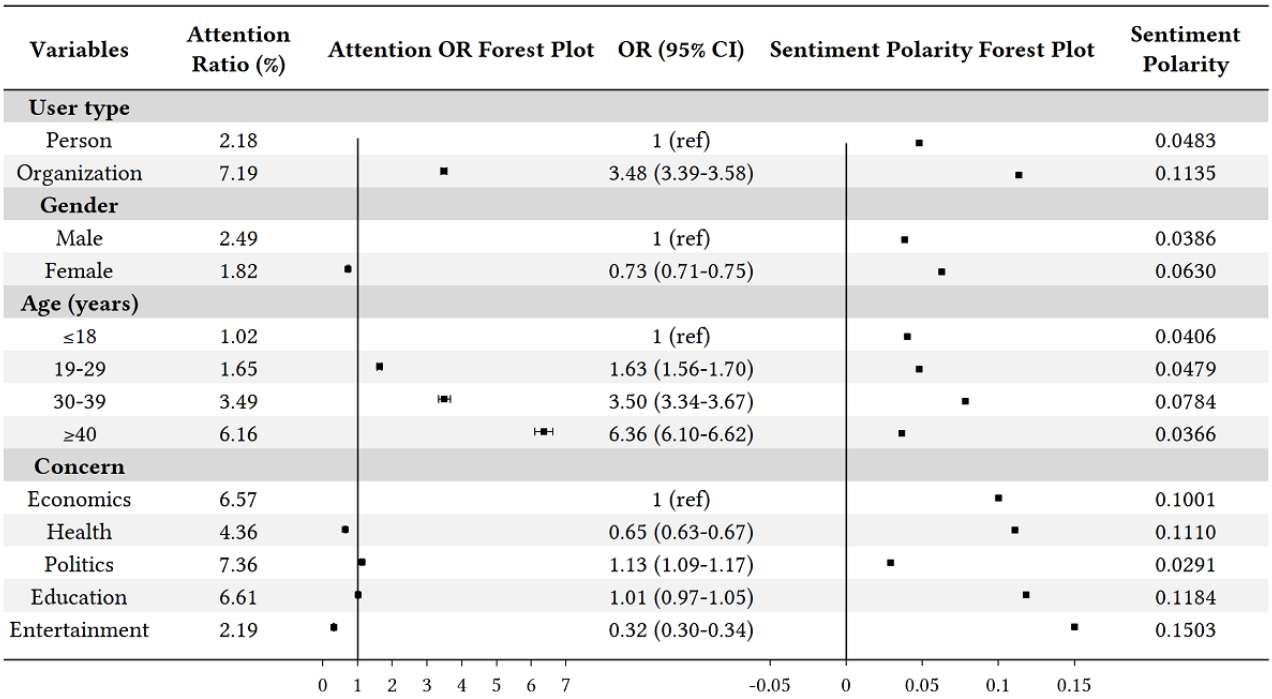
COVID-19 Concern and Sentiment Polarity Analysis

Univariate Analysis

The population characteristics in this study included four variables—user type, gender, age, and concern—on which we

first performed a univariate statistical analysis of COVID-19 concerns and sentiment polarities. The results are shown in Figure 4.

Figure 4. Univariate analysis of COVID-19 concerns and sentiment polarities among different population groups. OR: odds ratio.



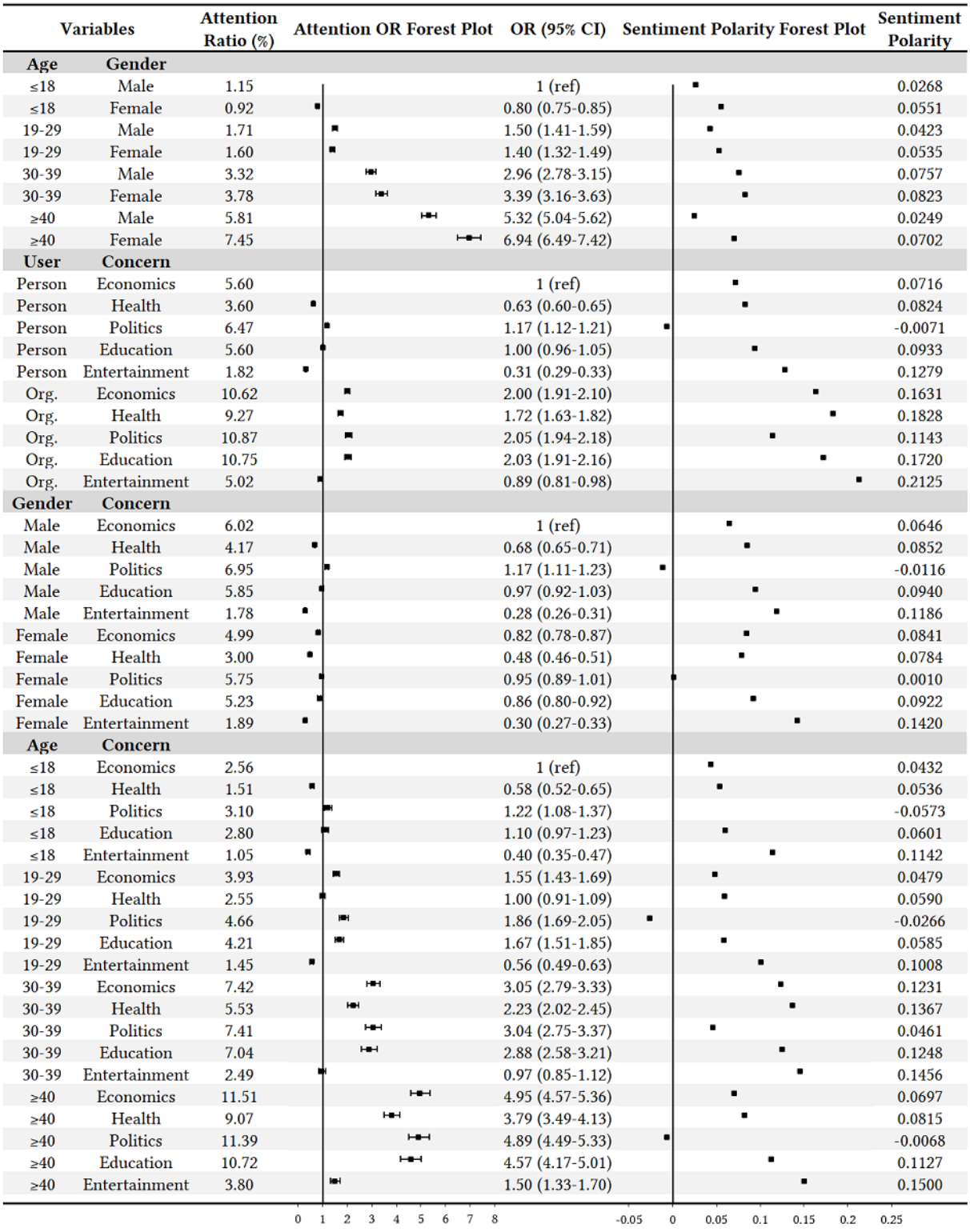
It can be seen that the organizations' attention ratio (7.19%) to COVID-19 was significantly higher than that of individuals (2.18%), and the attention OR of organizations was 3.48 (95% CI 3.39-3.58) compared with individuals. Moreover, organizations' sentiment polarity (0.1135) was more positive than that of individuals (0.0483). The COVID-19 attention ratio of females (1.82%) was a bit lower than that of males (2.49%), with an attention OR of 0.73 (95% CI 0.71-0.75). Meanwhile, females were more positive than males, and the sentiment polarities were 0.0630 and 0.0386 for females and males, respectively. In addition, COVID-19 attention increased significantly with age. Among the four age groups, the attention ORs of the groups that were 19 to 29 years, 30 to 39 years , and 40 years or older were 1.63 (95% CI 1.56-1.70), 3.50 (95% CI 3.34-3.67) and 6.36 (95% CI 6.10-6.62), respectively, in comparison with the group that was 18 years or less, which implies that older people are more concerned about COVID-19. The group that was 40 years or older was less positive than other age groups, with a sentiment polarity of 0.0366. For the concern variable, the COVID-19 attention ratios for politics (7.36%), education (6.61%), and economics (6.57%) were relatively high, followed by health (4.36%) and entertainment (2.19%). The sentiment polarity of political topics (0.0291) was the lowest among these topics, followed by economic (0.1001), health (0.1110), education (0.1184), and entertainment (0.1503) topics.

In general, these data indicate that organizations, as compared to individuals; males, as compared to females; and older people, as compared to young people, are more concerned about the pandemic. In addition, these data indicate that people are more concerned about politics, education, and economics under COVID-19.

Bivariate Analysis

Furthermore, we performed a bivariate analysis on COVID-19 attention and sentiment polarity by crossing any two population characteristic variables, as shown in Figure 5.

Figure 5. Bivariate analysis of COVID-19 concerns and sentiment polarities among different population groups. OR: odds ratio; Org: organization.



Variables		Attention Ratio (%)	OR (95% CI)	Sentiment Polarity
Age	**Gender**			
≤18	Male	1.15	1 (ref)	0.0268
≤18	Female	0.92	0.80 (0.75-0.85)	0.0551
19-29	Male	1.71	1.50 (1.41-1.59)	0.0423
19-29	Female	1.60	1.40 (1.32-1.49)	0.0535
30-39	Male	3.32	2.96 (2.78-3.15)	0.0757
30-39	Female	3.78	3.39 (3.16-3.63)	0.0823
≥40	Male	5.81	5.32 (5.04-5.62)	0.0249
≥40	Female	7.45	6.94 (6.49-7.42)	0.0702
User	**Concern**			
Person	Economics	5.60	1 (ref)	0.0716
Person	Health	3.60	0.63 (0.60-0.65)	0.0824
Person	Politics	6.47	1.17 (1.12-1.21)	-0.0071
Person	Education	5.60	1.00 (0.96-1.05)	0.0933
Person	Entertainment	1.82	0.31 (0.29-0.33)	0.1279
Org.	Economics	10.62	2.00 (1.91-2.10)	0.1631
Org.	Health	9.27	1.72 (1.63-1.82)	0.1828
Org.	Politics	10.87	2.05 (1.94-2.18)	0.1143
Org.	Education	10.75	2.03 (1.91-2.16)	0.1720
Org.	Entertainment	5.02	0.89 (0.81-0.98)	0.2125
Gender	**Concern**			
Male	Economics	6.02	1 (ref)	0.0646
Male	Health	4.17	0.68 (0.65-0.71)	0.0852
Male	Politics	6.95	1.17 (1.11-1.23)	-0.0116
Male	Education	5.85	0.97 (0.92-1.03)	0.0940
Male	Entertainment	1.78	0.28 (0.26-0.31)	0.1186
Female	Economics	4.99	0.82 (0.78-0.87)	0.0841
Female	Health	3.00	0.48 (0.46-0.51)	0.0784
Female	Politics	5.75	0.95 (0.89-1.01)	0.0010
Female	Education	5.23	0.86 (0.80-0.92)	0.0922
Female	Entertainment	1.89	0.30 (0.27-0.33)	0.1420
Age	**Concern**			
≤18	Economics	2.56	1 (ref)	0.0432
≤18	Health	1.51	0.58 (0.52-0.65)	0.0536
≤18	Politics	3.10	1.22 (1.08-1.37)	-0.0573
≤18	Education	2.80	1.10 (0.97-1.23)	0.0601
≤18	Entertainment	1.05	0.40 (0.35-0.47)	0.1142
19-29	Economics	3.93	1.55 (1.43-1.69)	0.0479
19-29	Health	2.55	1.00 (0.91-1.09)	0.0590
19-29	Politics	4.66	1.86 (1.69-2.05)	-0.0266
19-29	Education	4.21	1.67 (1.51-1.85)	0.0585
19-29	Entertainment	1.45	0.56 (0.49-0.63)	0.1008
30-39	Economics	7.42	3.05 (2.79-3.33)	0.1231
30-39	Health	5.53	2.23 (2.02-2.45)	0.1367
30-39	Politics	7.41	3.04 (2.75-3.37)	0.0461
30-39	Education	7.04	2.88 (2.58-3.21)	0.1248
30-39	Entertainment	2.49	0.97 (0.85-1.12)	0.1456
≥40	Economics	11.51	4.95 (4.57-5.36)	0.0697
≥40	Health	9.07	3.79 (3.49-4.13)	0.0815
≥40	Politics	11.39	4.89 (4.49-5.33)	-0.0068
≥40	Education	10.72	4.57 (4.17-5.01)	0.1127
≥40	Entertainment	3.80	1.50 (1.33-1.70)	0.1500

It can be seen that many results are consistent with the univariate analysis in the previous section. For example, under the combination of age and gender variables, the attention ratios grew with age, both for males and females. Moreover, females were more positive than males in all age groups. Under the combination of user type and concern variables, the order of concerns for individuals is politics, education, economics, health, and entertainment, which is similar to the univariate results.

However, there are still some noteworthy differences. First, not all females of different ages paid less attention to COVID-19 than males, but as individuals got older, females became more concerned than males, with the highest attention ratio of 7.45% and OR of 6.94 (95% CI 6.49-7.42) in females 40 years or older. Second, males 40 years or older (0.0249) and 18 years or younger (0.0268) were the least positive among all population groups. Third, different from the univariate concern analysis, the order of concerns for groups 30 to 39 years and 40 years or

older changed to economics, politics, education, health, and entertainment.

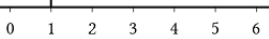
From the bivariate results, we can see that not all the population groups obeyed the same rules, but some of them presented worthy differences under multivariable conditions. We further conducted a deeper exploration in the following trivariate analysis.

Trivariate Analysis

In this part of the study, we crossed the three variables—gender, age, and concern—of population characteristics to study the COVID-19 responses, and a total of 40 combinations were produced, as shown in Figure 6. Since gender and age attributes did not exist in the organization group, this trivariate analysis only concentrated on individuals.

Figure 6. Trivariate analysis of COVID-19 concerns and sentiment polarities among different population groups. OR: odds ratio.

Variables			Attention Ratio (%)	Attention OR Forest Plot	OR (95% CI)	Sentiment Polarity Forest Plot	Sentiment Polarity
Gender	Age	Concern					
Male	≤18	Economics	2.65		1 (ref)		0.0361
Male	≤18	Health	1.75		0.66 (0.57-0.76)		0.0629
Male	≤18	Politics	3.38		1.29 (1.10-1.50)		-0.0658
Male	≤18	Education	3.17		1.20 (1.03-1.41)		0.0444
Male	≤18	Entertainment	1.06		0.39 (0.32-0.48)		0.0991
Male	19-29	Economics	4.12		1.58 (1.41-1.77)		0.0292
Male	19-29	Health	2.73		1.03 (0.91-1.17)		0.0613
Male	19-29	Politics	4.83		1.87 (1.64-2.13)		-0.0757
Male	19-29	Education	4.13		1.58 (1.38-1.82)		0.0402
Male	19-29	Entertainment	1.36		0.51 (0.43-0.60)		0.0434
Male	30-39	Economics	6.96		2.75 (2.45-3.09)		0.1194
Male	30-39	Health	5.40		2.10 (1.84-2.39)		0.1381
Male	30-39	Politics	6.61		2.60 (2.28-2.97)		0.0777
Male	30-39	Education	6.25		2.45 (2.12-2.83)		0.1368
Male	30-39	Entertainment	1.84		0.69 (0.57-0.83)		0.2053
Male	≥40	Economics	10.97		4.53 (4.08-5.03)		0.0653
Male	≥40	Health	8.54		3.43 (3.08-3.83)		0.0815
Male	≥40	Politics	11.62		4.83 (4.33-5.39)		-0.0082
Male	≥40	Education	10.18		4.16 (3.71-4.68)		0.1175
Male	≥40	Entertainment	3.54		1.35 (1.16-1.56)		0.1368
Female	≤18	Economics	2.46		0.93 (0.81-1.07)		0.0518
Female	≤18	Health	1.32		0.49 (0.43-0.57)		0.0443
Female	≤18	Politics	2.78		1.05 (0.89-1.25)		-0.0456
Female	≤18	Education	2.38		0.90 (0.75-1.07)		0.0840
Female	≤18	Entertainment	1.04		0.39 (0.31-0.47)		0.1315
Female	19-29	Economics	3.71		1.42 (1.26-1.59)		0.0712
Female	19-29	Health	2.40		0.90 (0.80-1.02)		0.0567
Female	19-29	Politics	4.47		1.72 (1.50-1.97)		0.0301
Female	19-29	Education	4.31		1.65 (1.44-1.90)		0.0785
Female	19-29	Entertainment	1.56		0.58 (0.49-0.70)		0.1675
Female	30-39	Economics	8.28		3.32 (2.92-3.78)		0.1289
Female	30-39	Health	5.70		2.22 (1.94-2.55)		0.1351
Female	30-39	Politics	9.12		3.69 (3.17-4.29)		-0.0031
Female	30-39	Education	8.52		3.42 (2.91-4.02)		0.1084
Female	30-39	Entertainment	4.27		1.64 (1.34-2.01)		0.0759
Female	≥40	Economics	13.36		5.67 (5.02-6.41)		0.0823
Female	≥40	Health	10.70		4.40 (3.86-5.02)		0.0815
Female	≥40	Politics	10.82		4.46 (3.90-5.10)		-0.0029
Female	≥40	Education	12.54		5.27 (4.54-6.12)		0.1001
Female	≥40	Entertainment	4.80		1.85 (1.49-2.30)		0.1876



Like in the bivariate analysis, there were some consistent results in the trivariate analysis. For example, the COVID-19 attention ratios increased with age, both for males and females in each topic of concern. Meanwhile, many detailed population differences were also clearly shown in these trivariate results. First, we can see that all of the groups presented different amounts of attention on the five topics of concern. In particular, females 40 years or older paid the greatest amount of attention to economic topics (OR 5.67, 95% CI 5.02-6.41), followed by education topics (OR 5.27, 95% CI 4.54-6.12). As a comparison, males in the same age group (ie, ≥40 years) had the highest concerns regarding political topics (OR 4.83, 95% CI 4.33-5.39), followed by economic (OR 4.53, 95% CI 4.08-5.03) and education (OR 4.16, 95% CI 3.71-4.68) topics. Second, the sentiment polarities of political topics were the lowest in all population groups, of which six had negative values. Lastly, the sentiment polarities of entertainment topics were always the highest among the five topics of concern across all population groups.

COVID-19 Emotion Analysis

We applied three different emotion models—Ekman's six basic emotions, Plutchik's eight primary emotions, and POMS six mood states—to perform emotion detection, both on the original

tweets and the COVID-19 tweets. The comparison results are shown in Figures 7 and 8. Figure 7 presents the mean intensity scores of the three emotion models, and Figure 8 shows the population distribution for each emotion from the models based on both original and COVID-19 tweets. As Ekman's six basic emotions (ie, anger, disgust, fear, joy, sadness, and surprise) are included in Plutchik's eight emotions, and these six common emotions had the same proportion rank in our experimental results, we only then analyzed Plutchik's and POMS emotions.

Figure 7. The mean intensity scores for the three emotion models. Scores range from 0 to 1 for each emotion. POMS: Profile of Mood States.

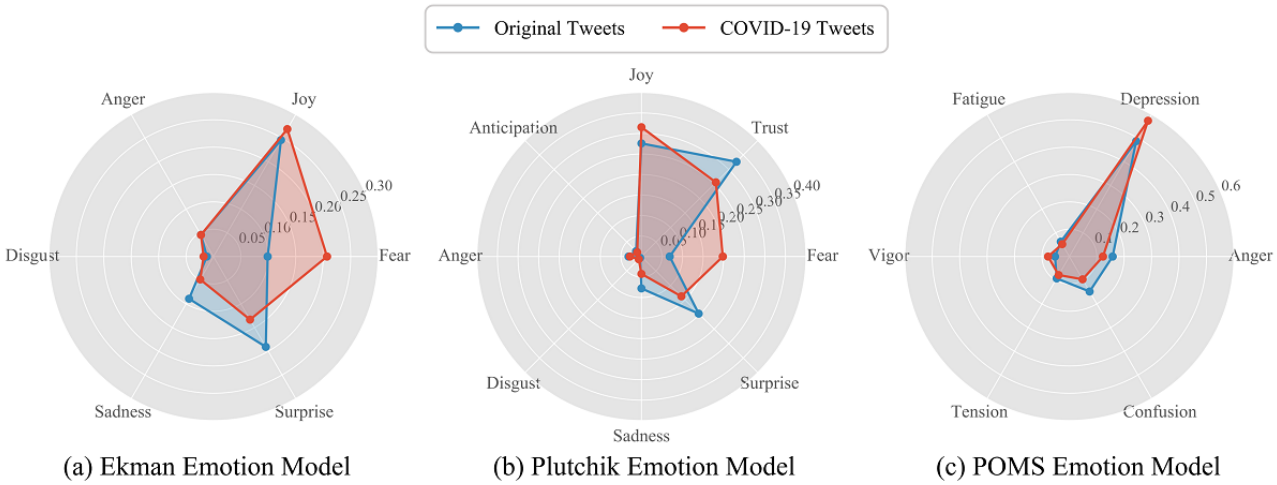
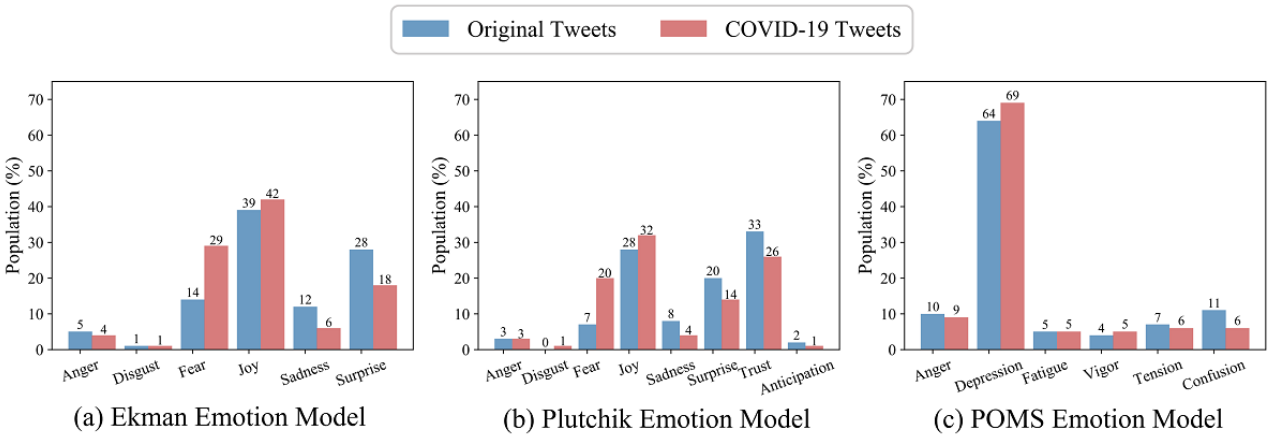


Figure 8. The population distributions of the three emotion models. POMS: Profile of Mood States.



In general, when Plutchik's emotion model was applied to the original tweets, *trust*, *joy*, and *surprise* were the highest emotions. When the model was applied to COVID-19 tweets, *fear* increased significantly, then *joy*, *trust*, and *fear* became the highest emotions. Meanwhile, when the POMS emotion model was applied to original tweets, *depression* was the most prominent emotion, and when applied to COVID-19 tweets, *depression* became even more prominent.

Afterward, we studied the differences in emotions considering the population characteristic attributes under COVID-19 by performing a chi-square test on each population attribute for each emotion. The results are shown in Multimedia Appendix 2. Figures 9 and 10 illustrate the emotion analysis by applying Plutchik's and POMS models to each population characteristic. We observed differences in emotions with respect to population variables, but among all the dominant emotions after applying Plutchik's and POMS models, *fear* and *depression* had significantly different scores and proportions in different populations. A further detailed statistical analysis was conducted on these two emotions (see Figure 11). We can see that organizations expressed more *fear* and *depression* than individuals, and females expressed less *fear* and *depression* than males. With increasing age, *fear* and *depression* increased significantly; in addition, people expressed more *fear* regarding political and health topics, and more *depression* regarding entertainment, economic, and political topics.

Figure 9. Plutchik emotion analysis on four population characteristics. Scores range from 0 to 1 for each emotion.

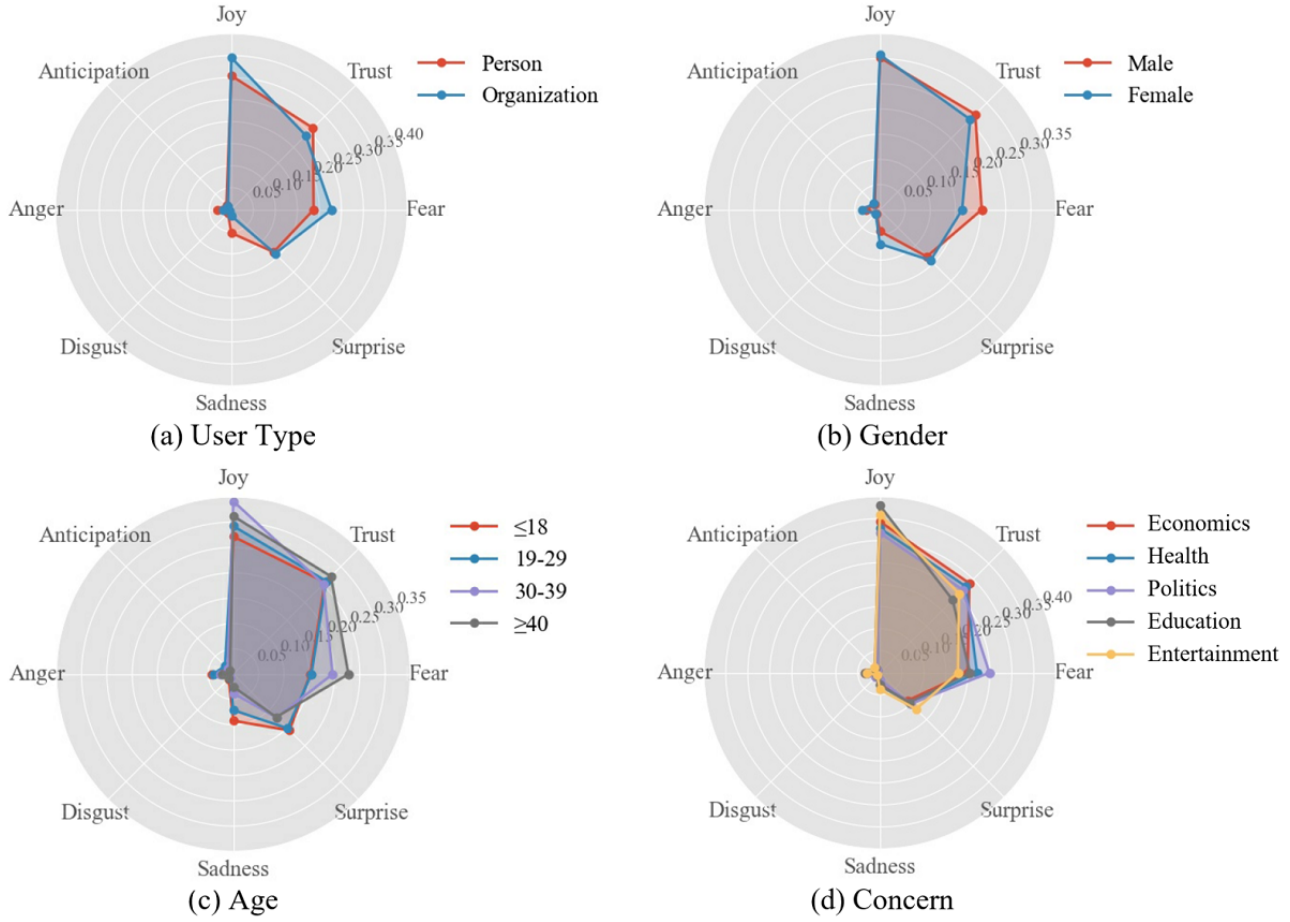


Figure 10. Profile of Mood States (POMS) emotion analysis on four population characteristics. Scores range from 0 to 1 for each emotion.

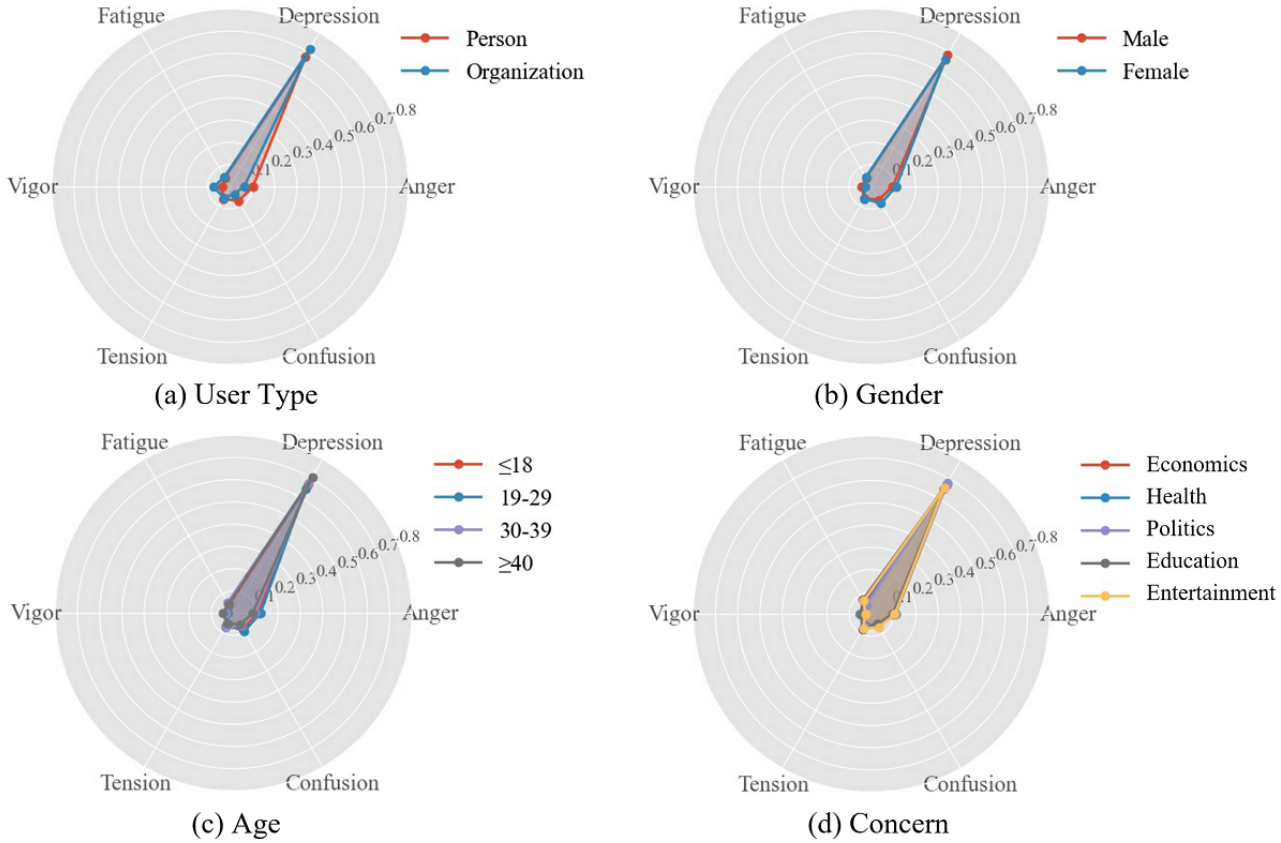


Figure 11. Statistical analysis of emotions related to COVID-19. OR: odds ratio.

Variables	Fear			Depression		
	Ratio (%)	Forest Plot	OR (95% CI)	Ratio (%)	Forest Plot	OR (95% CI)
User type						
Person	19.78		1 (ref)	68.93		1 (ref)
Organization	23.16		1.22 (1.15-1.30)	72.93		1.21 (1.15-1.28)
Gender						
Male	21.26		1 (ref)	69.87		1 (ref)
Female	16.86		0.75 (0.69-0.81)	67.06		0.88 (0.82-0.94)
Age (years)						
≤18	14.88		1 (ref)	65.47		1 (ref)
19-29	17.78		1.24 (1.10-1.39)	66.51		1.05 (0.96-1.15)
30-39	21.58		1.57 (1.39-1.79)	67.91		1.12 (1.01-1.24)
≥40	23.08		1.72 (1.53-1.92)	73.36		1.45 (1.33-1.59)
Concern						
Economics	20.95		1 (ref)	68.33		1 (ref)
Health	22.95		1.12 (1.04-1.22)	66.80		0.93 (0.87-1.00)
Politics	24.22		1.21 (1.11-1.31)	68.33		1.00 (0.93-1.07)
Education	21.38		1.03 (0.94-1.12)	65.33		0.87 (0.81-0.94)
Entertainment	21.43		1.03 (0.91-1.17)	68.97		1.03 (0.92-1.15)

In summary regarding the emotion analysis, it can be concluded that the emotions differed between original tweets and COVID-19 tweets, and they further differed among different population groups during the COVID-19 pandemic.

Discussion

Principal Findings

In this study, we analyzed a large amount of Twitter data collected from August 7 to 12, 2020, during the COVID-19 pandemic. In the overall analysis, the average sentiment polarity of COVID-19-related tweets posted by participants was less positive than that of the original tweets. In addition, the population groups under each variable (ie, user type, gender, age, and concern) all showed significant differences (P<.001) in response to COVID-19. In univariate analysis, organizations, as compared to individuals; males, as compared to females; and older people, as compared to young people were more concerned about the pandemic and had greater proportions of *fear* and *depression* emotions. In addition, the COVID-19 attention ratios of politics, education, and economics were relatively high, followed by health and entertainment, while the sentiment polarity of politics was the lowest, followed by economics, health, education, and entertainment.

Furthermore, the multivariate analyses showed more fine-grained and meaningful results. Among the findings, it is worth noting that not all female groups paid less attention to COVID-19 than male groups in the same age range, and not all groups' top concerns were the same. As age increased to above 30 years, females were gradually more concerned about COVID-19 than males. Moreover, females above 40 years of age were the group most concerned about COVID-19, and they were most concerned about economics and education. As a comparison, males in the same age group were most concerned about politics and economics. Males above 40 years of age and below 18 years of age were the least positive in sentiment. Among all the five topics of concern, the sentiment polarities of politics were the lowest in all population groups. These findings demonstrate that there exist population-level disparities in concerns and sentiments about COVID-19 in response to the pandemic during our research period.

We speculate that there are two reasons for the population-level differences. First, they are related to the concrete needs of specific age groups. For example, people older than 30 years of age may pay more attention to COVID-19 impacts on economics, while young people may concentrate more on education. Second, they are also related to the features of this novel coronavirus. Epidemiological studies have shown that the older population is more susceptible to COVID-19 and mortalities among this age group are higher than in other populations [13].

Limitations

The algorithm of demographic characteristic extraction used in this study is only capable of extracting three basic attributes: user type, gender, and age. Therefore, it is difficult for us to conduct a more detailed multivariable analysis compared with traditional questionnaire methods. In addition, the age range divisions were not fine-grained enough for COVID-19, especially for the group that was 40 years old or above, which covers a wide age range. To support the extraction of more attributes with finer granularity, we plan to optimize the current algorithm or seek new suitable and efficient algorithms for future studies.

Conclusions

Through large-scale Twitter data mining, this study revealed that salient disparities exist among population groups in terms of their concerns and sentiments regarding COVID-19-related issues. Therefore, it is suggested that government agencies and social organizations should devote specialized attention and support to each population group based on their varied concerns and sentiments experienced during the pandemic. The open source code developed in this study, which was publicly released via GitHub [22], can be easily employed to explore the evolution of population groups regarding their wants, needs, and thoughts during the pandemic for future follow-ups. It can also be repurposed for research and interventions used in combatting other public health emergencies, thanks to the efficient and economic nature of its operation.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (Grant Nos. 61876150 and 12026609) and the Science and Technology Program of the City of Xi'an (Grant Nos. 20YXYJ0009-12 and XA2020-RKXYJ-0105).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental information about the M3 (multimodal, multilingual, and multi-attribute) model.
[DOCX File , 32 KB - jmir_v23i3e26482_app1.docx]

Multimedia Appendix 2
Extended details on the data analyses.
[DOCX File , 31 KB - jmir_v23i3e26482_app2.docx]

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
2. WHO coronavirus disease (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2020-08-12]
3. Moreland A, Herlihy C, Tynan MA, Sunshine G, McCord RF, Hilton C, CDC Public Health Law Program, CDC COVID-19 Response Team, Mitigation Policy Analysis Unit. Timing of state and territorial COVID-19 stay-at-home orders and changes in population movement - United States, March 1-May 31, 2020. MMWR Morb Mortal Wkly Rep 2020 Sep 04;69(35):1198-1203 [FREE Full text] [doi: 10.15585/mmwr.mm6935a2] [Medline: 32881851]
4. Mayo Clinic Staff. COVID-19 quarantine, self-isolation and social distancing. Mayo Clinic. 2020. URL: https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/coronavirus-quarantine-and-isolation/art-20484503 [accessed 2020-08-12]
5. Impact of COVID-19 on people's livelihoods, their health and our food systems. World Health Organization. 2020 Oct 13. URL: https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people's-livelihoods-their-health-and-our-food-systems [accessed 2020-12-01]
6. Policy Brief: Education During COVID-19 and Beyond. New York, NY: United Nations; 2020 Aug. URL: https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/08/sg_policy_brief_covid-19_and_education_august_2020.pdf [accessed 2020-12-01]
7. United Nations Conference on Trade and Development. COVID-19 and E-Commerce: Impact on Businesses and Policy Responses. Geneva, Switzerland: United Nations; 2020 Nov 17. URL: https://unctad.org/system/files/official-document/dtlstict2020d12_en.pdf [accessed 2021-02-26]
8. United Nations Conference on Trade and Development. Impact of the Pandemic on Trade and Development: Transitioning to a New Normal. Geneva, Switzerland: United Nations; 2020. URL: https://unctad.org/system/files/official-document/osg2020d1_en.pdf [accessed 2021-02-26]
9. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of twitter trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
10. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]

11. Hung M, Lauren E, Hon ES, Birmingham WC, Xu J, Su S, et al. Social network analysis of COVID-19 sentiments: Application of artificial intelligence. J Med Internet Res 2020 Aug 18;22(8):e22590 [FREE Full text] [doi: 10.2196/22590] [Medline: 32750001]
12. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
13. Yuki K, Fujiogi M, Koutsogiannaki S. COVID-19 pathophysiology: A review. Clin Immunol 2020 Jun;215:108427 [FREE Full text] [doi: 10.1016/j.clim.2020.108427] [Medline: 32325252]
14. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
15. Liu N, Zhang F, Wei C, Jia Y, Shang Z, Sun L, et al. Prevalence and predictors of PTSS during COVID-19 outbreak in China hardest-hit areas: Gender differences matter. Psychiatry Res 2020 May;287:112921 [FREE Full text] [doi: 10.1016/j.psychres.2020.112921] [Medline: 32240896]
16. Lu W, Wang H, Lin Y, Li L. Psychological status of medical workforce during the COVID-19 pandemic: A cross-sectional study. Psychiatry Res 2020 Jun;288:112936 [FREE Full text] [doi: 10.1016/j.psychres.2020.112936] [Medline: 32276196]
17. Nelson LM, Simard JF, Oluyomi A, Nava V, Rosas LG, Bondy M, et al. US public concerns about the COVID-19 pandemic from results of a survey given via social media. JAMA Intern Med 2020 Jul 01;180(7):1020-1022 [FREE Full text] [doi: 10.1001/jamainternmed.2020.1369] [Medline: 32259192]
18. Groarke JM, Berry E, Graham-Wisener L, McKenna-Plumley PE, McGlinchey E, Armour C. Loneliness in the UK during the COVID-19 pandemic: Cross-sectional results from the COVID-19 Psychological Wellbeing Study. PLoS One 2020;15(9):e0239698 [FREE Full text] [doi: 10.1371/journal.pone.0239698] [Medline: 32970764]
19. Azlan AA, Hamzah MR, Sern TJ, Ayub SH, Mohamad E. Public knowledge, attitudes and practices towards COVID-19: A cross-sectional study in Malaysia. PLoS One 2020;15(5):e0233668 [FREE Full text] [doi: 10.1371/journal.pone.0233668] [Medline: 32437434]
20. Ahmad AR, Murad HR. The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: Online questionnaire study. J Med Internet Res 2020 May 19;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]
21. van de Mortel TF. Faking it: Social desirability response bias in self-report research. Aust J Adv Nurs 2008;25(4):40-48 [FREE Full text]
22. Understanding concerns, sentiments and disparities of population groups during the COVID-19 pandemic. GitHub. 2020. URL: https://github.com/cyzhang87/EmulatedQuestionnaireOnTwitter [accessed 2021-02-26]
23. Sampled stream v1. Twitter Developer. 2020. URL: https://developer.twitter.com/en/docs/labs/sampled-stream/overview [accessed 2020-02-01]
24. Sampled stream. Twitter Developer. 2020. URL: https://developer.twitter.com/en/docs/twitter-api/tweets/sampled-stream/introduction [accessed 2020-10-01]
25. COVID-19 stream. Twitter Developer. 2020. URL: https://developer.twitter.com/en/docs/labs/covid19-stream/filtering-rules [accessed 2020-07-01]
26. Wang Z, Hale S, Adelani DI, Grabowicz P, Hartman T, Flöck F, et al. Demographic inference and representative population estimates from multilingual social media data. In: Proceedings of the World Wide Web Conference (WWW '19). New York, NY: Association for Computing Machinery; 2019 May Presented at: The World Wide Web Conference (WWW '19); May 13-17, 2019; San Francisco, CA p. 2056-2067. [doi: 10.1145/3308558.3313684]
27. Rothe R, Timofte R, Van Gool L. Deep expectation of real and apparent age from a single image without facial landmarks. Int J Comput Vis 2016 Aug 10;126(2-4):144-157. [doi: 10.1007/s11263-016-0940-3]
28. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York, NY: IEEE; 2017 Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 21-26, 2017; Honolulu, HI p. 2261-2269. [doi: 10.1109/cvpr.2017.243]
29. Hashimzade N, Myles G, Black J. A Dictionary of Economics. 5th edition. Oxford, UK: Oxford University Press; 2017.
30. Economics A-Z terms beginning with A. The Economist. 2020. URL: https://www.economist.com/economics-a-to-z [accessed 2020-06-01]
31. Brown GW, McLean I, McMillan A. A Concise Oxford Dictionary of Politics and International Relations. 4th edition. Oxford, UK: Oxford University Press; 2018.
32. Hutto CJ, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the 8th International Conference on Weblogs and Social Media (ICWSM-14). 2014 Presented at: 8th International Conference on Weblogs and Social Media (ICWSM-14); June 1-4, 2014; Ann Arbor, MI p. 1-10 URL: http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf
33. Colneric N, Demsar J. Emotion recognition on Twitter: Comparative study and training a unison model. IEEE Trans Affect Comput 2020 Jul 1;11(3):433-446. [doi: 10.1109/taffc.2018.2807817]

34. Ekman P. An argument for basic emotions. Cogn Emot 1992;6(3-4):169-200. [doi: 10.1080/02699939208411068]
35. Plutchik R. A general psychoevolutionary theory of emotion. In: Plutchik R, Kellerman H, editors. Emotion: Theory, Research, and Experience, Volume 1: Theories of Emotion. New York, NY: Academic Press; 1980:3-33.
36. McNair DM, Lorr M, Droppleman LF. Manual for the Profile of Mood States. San Diego, CA: Educational and Industrial Testing Services; 1971.

Abbreviations

DenseNet: dense convolutional network
M3: multimodal, multilingual, and multi-attribute
OR: odds ratio
POMS: Profile of Mood States
VADER: Valence Aware Dictionary and Emotional Reasoner

Edited by R Kukafka, C Basch; submitted 14.12.20; peer-reviewed by L Sheets, K Acquaviva; comments to author 15.01.21; revised version received 05.02.21; accepted 18.02.21; published 05.03.21.

Please cite as:
Zhang C, Xu S, Li Z, Hu S
Understanding Concerns, Sentiments, and Disparities Among Population Groups During the COVID-19 Pandemic Via Twitter Data Mining: Large-scale Cross-sectional Study
J Med Internet Res 2021;23(3):e26482
URL: https://www.jmir.org/2021/3/e26482
doi:10.2196/26482
PMID:33617460

Original Paper

Comparing Public Perceptions and Preventive Behaviors During the Early Phase of the COVID-19 Pandemic in Hong Kong and the United Kingdom: Cross-sectional Survey Study

Leigh Bowman[1], PhD; Kin On Kwok[2,3,4], PhD; Rozlyn Redd[5], PhD; Yuanyuan Yi[2], MPH; Helen Ward[1,5], PhD; Wan In Wei[2], MPhil; Christina Atchison[5], PhD; Samuel Yeung-Shan Wong[2], MD

[1]MRC Centre for Global Infectious Disease Analysis and Abdul Latif Jameel Institute for Disease and Emergency Analytics (J-IDEA), School of Public Health, Imperial College, London, United Kingdom

[2]JC School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, Hong Kong

[3]Stanley Ho Centre for Emerging Infectious Diseases, The Chinese University of Hong Kong, Hong Kong, Hong Kong

[4]Shenzhen Research Institute, The Chinese University of Hong Kong, Hong Kong, Hong Kong

[5]Patient Experience Research Centre, School of Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Kin On Kwok, PhD
JC School of Public Health and Primary Care
The Chinese University of Hong Kong
Room 419, 4/F, JC School of Public Health and Primary Care Building, Prince of Wales Hospital
Shatin, New Territories
Hong Kong
Hong Kong
Phone: 852 22528405
Email: kkokwok@cuhk.edu.hk

Abstract

Background: Given the public health responses to previous respiratory disease pandemics, and in the absence of treatments and vaccines, the mitigation of the COVID-19 pandemic relies on population engagement in nonpharmaceutical interventions. This engagement is largely driven by risk perception, anxiety levels, and knowledge, as well as by historical exposure to disease outbreaks, government responses, and cultural factors.

Objective: The aim of this study is to compare psychobehavioral responses in Hong Kong and the United Kingdom during the early phase of the COVID-19 pandemic.

Methods: Comparable cross-sectional surveys were administered to adults in Hong Kong and the United Kingdom during the early phase of the epidemic in each setting. Explanatory variables included demographics, risk perception, knowledge of COVID-19, anxiety level, and preventive behaviors. Responses were weighted according to census data. Logistic regression models, including effect modification to quantify setting differences, were used to assess the association between the explanatory variables and the adoption of social distancing measures.

Results: Data from 3431 complete responses (Hong Kong, 1663; United Kingdom, 1768) were analyzed. Perceived severity of symptoms differed by setting, with weighted percentages of 96.8% for Hong Kong (1621/1663) and 19.9% for the United Kingdom (366/1768). A large proportion of respondents were abnormally or borderline anxious (Hong Kong: 1077/1603, 60.0%; United Kingdom: 812/1768, 46.5%) and regarded direct contact with infected individuals as the transmission route of COVID-19 (Hong Kong: 94.0%-98.5%; United Kingdom: 69.2%-93.5%; all percentages weighted), with Hong Kong identifying additional routes. Hong Kong reported high levels of adoption of various social distancing measures (Hong Kong: 32.6%-93.7%; United Kingdom: 17.6%-59.0%) and mask-wearing (Hong Kong: 98.8% (1647/1663); United Kingdom: 3.1% (53/1768)). The impact of perceived severity of symptoms and perceived ease of transmission of COVID-19 on the adoption of social distancing measures varied by setting. In Hong Kong, these factors had no impact, whereas in the United Kingdom, those who perceived their symptom severity as “high” were more likely to adopt social distancing (adjusted odds ratios [aORs] 1.58-3.01), and those who perceived transmission as “easy” were prone to adopt both general social distancing (aOR 2.00, 95% CI 1.57-2.55) and contact avoidance (aOR 1.80, 95% CI 1.41-2.30). The impact of anxiety on adopting social distancing did not vary by setting.

Conclusions: Our results suggest that health officials should ascertain baseline levels of risk perception and knowledge in populations, as well as prior sensitization to infectious disease outbreaks, during the development of mitigation strategies. Risk should be communicated through suitable media channels—and trust should be maintained—while early intervention remains the cornerstone of effective outbreak response.

(J Med Internet Res 2021;23(3):e23231) doi:10.2196/23231

KEYWORDS

COVID-19; novel coronavirus; pandemic; behavioural response; risk perceptions; anxiety; comparative; Hong Kong; United Kingdom

Introduction

In December 2019, a novel coronavirus, SARS-CoV-2, emerged in Wuhan, Hubei Province, China, and spread rapidly worldwide, forming the second pandemic of the 21st century [1]. The progression of the disease has varied by region. As of August 2, 2020, there have been at least 17 million cases of COVID-19 and over 670,000 deaths globally [2].

Prior to the availability of effective treatments and vaccines, strategies to mitigate the impact of the pandemic have been primarily nonpharmaceutical [3], mainly focusing on public health promotion of using simple but effective preventive measures [4,5]. Many important control strategies currently promoted by governments require public participation, either through direct adoption of preventative behaviors, such as handwashing or wearing face masks, or through compliance with social distancing policies, such as recommendations to avoid public transport and mass gatherings.

Previous studies of the severe acute respiratory syndrome and influenza pandemics showed that governments should account for risk perception and anxiety when promoting preventative measures. There is evidence that higher perceived risk of infection is associated with increased adoption of precautionary measures [6,7], while increased anxiety has also been shown to increase the likelihood that people will engage in protective behaviors [8]. Moreover, longitudinal data suggest that these perceptions, behaviors and anxieties change with context and over time as uncertainty about disease severity decreases and knowledge of transmission increases [9].

During the current COVID-19 pandemic, researchers have examined public risk perceptions and knowledge in various countries, including Finland [10], Israel [11], Italy [12], Nigeria [13], the United States [14,15], South Korea [16] and Vietnam [17]. However, only a few studies have identified the factors associated with greater adoption of preventative measures or how these associations vary by context. In Hong Kong, both greater understanding of COVID-19 and increased anxiety were associated with greater adoption of social distancing behaviors [5], whereas in the United Kingdom, there was a significant socioeconomic gradient in the ability to adopt and comply with social distancing measures, specifically the ability to work from home and the ability to self-isolate [18].

This initial evidence that there is variation across context in affective responses, risk perceptions, and the impact of sociodemographic factors on the uptake of preventative behaviors has significant implications when tailoring policies. To elucidate these relationships, a more thorough comparative analysis is required. However, studies in different countries often use different metrics to measure the same behavior, which can lead to difficulty when interpreting the significance of heterogeneous contexts.

In this study, we examined and compared public perception and adoption of preventive behaviors in response to the early phase of the COVID-19 pandemic in two different settings: Hong Kong and the United Kingdom. We further investigated the factors associated with greater adoption of different types of social distancing measures. Our results have immediate implications on how health officials plan and communicate strategies to mitigate the ongoing COVID-19 pandemic to communities.

Methods

Study Design and Recruitment

In Hong Kong and the United Kingdom, cross-sectional surveys were conducted during the early phase of the COVID-19 pandemic, when limited government-level interventions were in place [5,18]. The survey period in Hong Kong was from January 24 to February 13, 2020, and that in the United Kingdom was from March 17 to 18, 2020. In Hong Kong, the first laboratory-confirmed case of COVID-19 was reported on January 23, 2020, and the number of cases rose to 53 by February 13, 2020 [19]; meanwhile, in the United Kingdom, the first two laboratory-confirmed cases were reported on January 31, 2020, and the number of cases rose to 2626 by March 18, 2020 [20].

In Hong Kong, all 452 district councilors were invited to distribute an open web-based survey by sharing a survey link and promotion messages on their webpages, social media platforms, or any channels which they usually used to convey information to their targeted residents. Individuals aged ≥18 years who understood Chinese and lived in Hong Kong were eligible to participate [5]. Respondents were compensated with HK $10 (US $1.29) in the form of a cash coupon. In the United Kingdom, the web-based survey was not open (users were required to log in) and was administered by YouGov, a market research company, to members of its panel of ≥800,000 individuals (aged ≥18 years) as part of their omnibus survey [21]. Our UK sample was obtained through a nonprobabilistic active sampling method, and emails were sent to randomly selected individuals with particular characteristics to match the proportions of people with those characteristics in the 2011 UK

census. No incentive was involved in the UK survey. More details of the survey design are described elsewhere [5,18].

Study Instruments

The study instruments are freely available on the web (Hong Kong: [22]; United Kingdom: [23]). The UK questionnaire was adapted from the Hong Kong version, with feedback from 20 members of the public (of different backgrounds) to improve its relevance and usability in the UK context. This process led to some discrepancies in the questions or answer choices; however, the two questionnaires were largely similar.

Sociodemographic variables included age, sex, educational attainment, and employment status. Anxiety level was measured using the Hospital Anxiety and Depression scale–Anxiety (HASD-A) (0-7=normal; 8-10=borderline abnormal; 11-21=abnormal) [24]. Risk perception toward COVID-19 was measured by perceived severity of symptoms if the respondent was infected with COVID-19 (Hong Kong, question 35; United Kingdom, question GIC_Q29). Knowledge of COVID-19 was assessed by asking whether COVID-19 could be transmitted through various routes (Hong Kong, question 45; United Kingdom, question GIC_Q33), including direct human exposure (eg, physical contact or a face-to-face conversation with someone who is infected with SARS-CoV-2 with or without symptoms) and other types of exposure (eg, visiting wet markets or consumption of wild animal meat). From knowledge of COVID-19, perceived ease of transmission was regarded as "easy" if the virus was deemed to be transmitted through face-to face conversation with asymptomatic infectees and as "difficult" otherwise. Respondents were also asked about the sources from which they retrieved information about COVID-19 and their perceived reliability of these sources (Hong Kong, questions 40 and 43; United Kingdom, questions GIC_Q30 and GIC_Q32). In addition, they were asked about the adoption of preventative behaviors to prevent the transmission of COVID-19 (Hong Kong, question 46; United Kingdom, questions GIC_Q34a and GIC_Q34b). Three types of preventative measures were considered: personal hygiene, social distancing, and travel avoidance.

Data Analysis

Descriptive statistics for all variables present the number of respondents and the raw or weighted percentages. In this manuscript, weighted percentages were used for description except for demographics. The responding samples were weighted to be representative of the United Kingdom (2011 census [25]) and Hong Kong (2016 by-census [26]) adult populations using the raking method [27]. Each data point was given a weight so that the marginal proportions of the demographics in the survey (age, sex, region, education level, and [United Kingdom only] social grade) were similar to those in the census. Chi-square goodness-of-fit tests were used for comparing characteristics across settings. Multivariate logistic regression models were used to identify sociodemographic and psychosocial factors associated with the adoption of three types of social distancing: (1) general measures, specified by avoiding crowded places, social events and going out; (2) contact measures, specified by avoiding contact with individuals who had fever or respiratory symptoms and who had been to affected areas recently; and (3) work measures, specified by avoiding going to work.

Common and comparable sociodemographic factors considered in separate analytical studies [5,18] were included in this comparative analysis. These factors were considered as confounders in the association between psychosocial factors (including anxiety level, perceived severity, and perceived ease of transmission) and adoption of social distancing measures. Further, these associations (between each aforementioned psychological factor and each type of social distancing measure) were considered a priori to be affected by setting. Therefore, we examined the effect modifications due to setting using interaction terms in the baseline models, which can be interpreted as the difference in the estimated effects of psychosocial factors on adopting social distancing measures due to different settings. Adjusted odds ratios (aORs) and 95% confidence intervals were estimated. Associations with $P<.05$ in the adjusted analyses were considered to be statistically significant. Analyses were conducted in R, version 3.6.3 (the R Project) and STATA, version 11 (StataCorp LLC).

Ethical Approval

The study was approved by the Imperial College London Research Ethics Committee (reference number: 20IC5861) and the Survey and Behavioral Research Ethics Committee of The Chinese University of Hong Kong (reference number: SBRE-19-625).

Results

Survey Responses

In Hong Kong, there were initially 2478 clicks on the survey link. After removing 763 cases with missing demographics and 52 cases with ambiguous responses on the perceived ease of transmission, 1663 complete cases were included in the analysis. In the United Kingdom, 2500 individuals were approached, and the response rate was 84.3% (2108/2500). After excluding cases with missing demographics or perceived severity and cases with ambiguous responses on the perceived ease of transmission, 1768 cases were included in the analysis.

Demographic Differences

There were significant differences in the sociodemographic characteristics of the study respondents between the two settings. Hong Kong respondents were younger, with 26.0% (433/1663) aged 18-24 years, compared with 9.4% (166/1768) for the United Kingdom ($P<.001$) (Table 1).

Table 1. Characteristics of the study respondents in the United Kingdom and Hong Kong (all *P* values <.001 as determined by chi-square goodness-of-fit test).

Characteristics	United Kingdom (n=1768)			Hong Kong (n=1663)		
	n	% (unweighted)	% (weighted)	n	% (unweighted)	% (weighted)
Age (years)						
18-24	166	9.4	9.9	433	26.0	17.0
25-34	243	13.7	14.3	535	32.2	23.5
35-44	335	18.9	19.5	370	22.2	23.9
45-54	300	17.0	17.7	193	11.6	22.2
≥55	724	41.0	38.6	132	7.9	13.4
Sex						
Female	936	52.9	51.8	1141	68.6	57.1
Male	832	47.1	48.2	522	31.4	42.9
Education attainment						
No formal qualification/lower secondary or below	100	5.7	5.5	53	3.2	9.9
Secondary level qualification/higher secondary	738	41.7	43.2	292	17.6	32.5
Postsecondary but below degree	334	18.9	18.3	267	16.1	16.2
Degree or above	596	33.7	32.9	1051	63.2	41.5
Employment status						
Employer/employee	1025	58.0	59.6	1135	68.3	66.0
Full-time student	90	5.1	5.3	278	16.7	12.5
Unemployed/not working	172	9.7	10.4	206	12.4	17.2
Retired	481	27.2	24.7	44	2.6	4.3
Perceived severity[a]						
Level 1	96	5.4	5.2	1071	64.4	65.0
Level 2	270	15.3	14.7	550	33.1	31.8
Level 3	1058	59.8	60.2	32	1.9	2.2
Level 4	320	18.1	18.5	7	0.4	0.7
Level 5	24	1.4	1.4	3	0.2	0.3
Worry about COVID-19						
Very worried	536	30.3	30.1	852	51.2	49.4
Fairly worried	858	48.5	48.4	723	43.5	43.2
Neutral/don’t know	5	0.3	0.3	40	2.4	3.2
Not very worried	295	16.7	16.9	1	0.1	0.1
Not at all worried	74	4.2	4.3	47	2.8	4.1
Anxiety level						
Normal	956	54.1	53.5	586	35.2	40.0
Borderline abnormal	336	19.0	19.4	512	30.8	27.3
Abnormal	476	26.9	27.1	565	34.0	32.7

[a]Level 1=very serious (Hong Kong)/life-threatening (United Kingdom); Level 2=serious (Hong Kong)/severe (eg, may need care and treatment in hospital) (United Kingdom); Level 3=neutral (Hong Kong)/moderate (eg, may need self-care and rest in bed) (United Kingdom); Level 4=not serious (Hong Kong)/mild (eg, can go about daily tasks normally) (United Kingdom); Level 5=not serious at all (Hong Kong)/no symptoms (United Kingdom).

The Hong Kong sample contained a greater proportion of women (Hong Kong: 1141/1663, 68.6%, vs United Kingdom: 936/1768, 52.9%; *P*<.001), and respondents educated to university degree level or above (Hong Kong: 1051/1663, 63.2%, vs United Kingdom: 596/1768, 33.7%; *P*<.001). Employment status reflected the age structure of the respondents in each setting, with a greater proportion of UK respondents in the retired category (Hong Kong: 44/1663, 2.6%, vs United Kingdom: 481/1768, 27.2%; *P*<.001) (Table 1).

Perceptions and Beliefs

Higher perceived severity of COVID-19 was observed among Hong Kong respondents, with 96.8% (1621/1663) rating the symptoms of COVID-19 infection as serious or very serious compared with only 19.9% (366/1768) of the UK respondents. In terms of levels of concern, 92.6% (1575/1663) of the Hong Kong sample responded that they felt very or fairly worried, compared with 78.5% (1394/1768) of the UK sample. The HADS-A scores reflected similar trends, with 60.0% (1077/1663) of the Hong Kong sample recording an abnormal or borderline abnormal result, compared with 46.5% (812/1768) of the UK sample (Table 1).

Knowledge and Information Sources

The majority of respondents regarded direct contact with infected individuals (Hong Kong: 94.0%-98.5%; United Kingdom: 69.2%-93.5%) or virus-contaminated environments (Hong Kong: 1594/1663, 96.3%; United Kingdom: 1411/1768, 79.5%) as the primary means of virus transmission (Table 2). However, respondents from Hong Kong identified a far broader scope of transmission routes. A much larger proportion of Hong Kong respondents regarded wild animal meat (Hong Kong: 1546/1663, 93.4%; United Kingdom: 199/1768, 11.3%), wet markets (Hong Kong: 1342/1663, 81.1%; United Kingdom: 374/1768, 21.5%), imported seafood (Hong Kong: 1199/1663, 70.9%; United Kingdom: 258/1768, 14.8%) and imported goods (Hong Kong: 1101/1663, 66.6%; United Kingdom: 209/1768, 12.1%) as potential exposure sources than their UK counterparts. There was also significant variation across use and reliability of information sources (Table S1 in Multimedia Appendix 1). The majority of respondents deemed health professionals to be reliable (>80% in both Hong Kong and the United Kingdom); however, few could access them (Hong Kong: 86/1663, 4.8%; United Kingdom: 202/1768, 11.5%). In addition, most UK respondents (1602/1768, 90.7%) considered official websites to be reliable, compared to 15.6% (260/1663) among Hong Kong respondents at the beginning of the pandemic.

Table 2. Knowledge of COVID-19 transmission.

"Are the following transmission routes of COVID-19?"	Respondents answering "yes"					
	United Kingdom (n=1768)			Hong Kong (n=1663)		
	n	% (unweighted)	% (weighted)	n	% (unweighted)	% (weighted)
Contact						
Face-to-face conversation (no physical contact) with someone who has SARS-CoV-2 but no symptoms	1234	69.8	69.2	1564	94.0	94.0
Face-to-face conversation (no physical contact) with someone who has SARS-CoV-2 with symptoms	1398	79.1	78.7	1616	97.2	96.8
Physical contact with someone who has SARS-CoV-2 but no symptoms	1580	89.4	89.0	1635	98.3	98.1
Physical contact with someone with SARS-CoV-2 who has symptoms	1657	93.7	93.5	1644	98.9	98.5
Transmission mode						
Droplets	N/A[a]	N/A	N/A	1649	99.2	99.2
Aerosol when infected people cough or sneeze	N/A	N/A	N/A	1478	88.9	91.2
Being in close contact (ie, within 2 meters) with someone who has SARS-CoV-2 when they cough or sneeze	1604	90.7	90.4	N/A	N/A	N/A
Being further away (ie, further than 2 meters) from someone who has SARS-CoV-2 when they cough or sneeze	615	34.8	34.8	N/A	N/A	N/A
Others						
Contact with virus-contaminated environment	1411	79.8	79.5	1594	95.9	96.3
Consumption of wild animal meat	199	11.3	11.3	1546	93.0	93.4
Visiting a wet market	374	21.2	21.5	1342	80.7	81.1
Consumption of seafood imported from specific regions[b]	258	14.6	14.8	1199	72.1	70.9
Consumption/use of products imported from specific regions[b]	209	11.8	12.1	1101	66.2	66.6

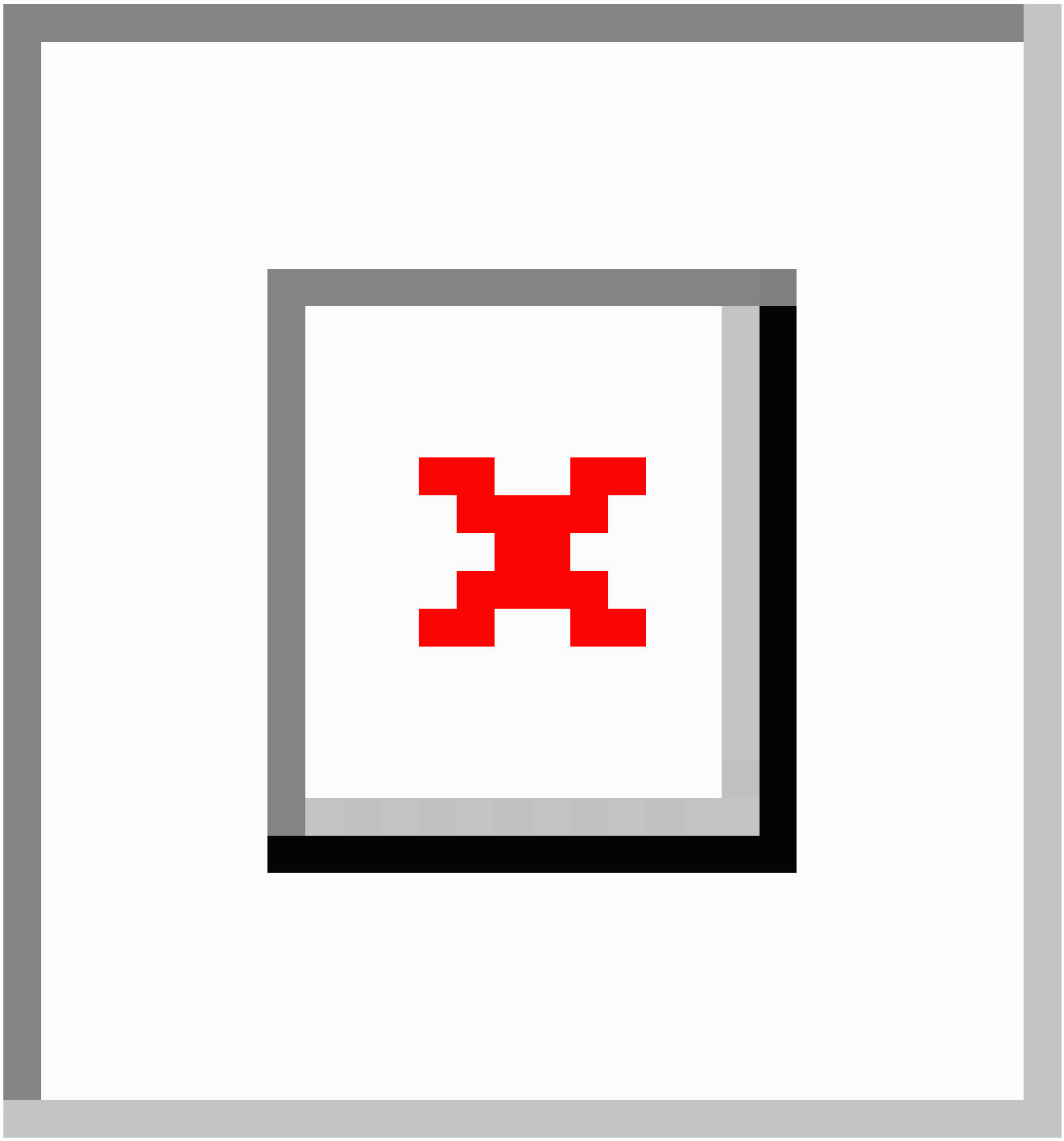
[a]N/A: not applicable.

[b]Specific regions refer to China (United Kingdom)/Wuhan (Hong Kong).

Adoption of Social Distancing Measures

There were variations in the weighted proportions of Hong Kong and the UK respondents who adopted precautionary measures against COVID-19 (Figure 1; Table S2 in Multimedia Appendix 1). Hong Kong respondents reported higher levels of adoption across all social distancing and personal hygiene measures. In particular, 98.8% (1647/1663) of Hong Kong respondents reported wearing a face mask, compared to 3.1% (53/1768) among the UK respondents. General measures were adopted by 63.1%-87.2% and 37.8%-59.0% of respondents in Hong Kong and the United Kingdom, respectively. Contact measures were adopted by 83.8%-93.7% and 33.7%-50.1% of respondents in Hong Kong and the United Kingdom, respectively. Work measures were reported by 32.6% (402/1135) and 22.5% (231/1025) of respondents in Hong Kong and the United Kingdom, respectively.

Figure 1. Adoption of precautionary measures against COVID-19. “Affected areas” refers to China (Hong Kong)/affected areas in the world (United Kingdom); “Specific regions in a limited period” refers to Wuhan in the past one month (Hong Kong)/affected areas in the past 14 days (United Kingdom). The “Going to work” category only included respondents who were employees or employers (n=2160), and the “Going to school/letting your children go to school” category only included respondents who were full-time students or had at least one child (n=1239).



Sociodemographic factors were associated with the three social distancing measures (Table S3 in Multimedia Appendix 1; Table 3). The UK respondents were significantly less likely than their Hong Kong counterparts to adopt social distancing measures (Table S3 in Multimedia Appendix 1, OR 0.08-0.53, *P*<.001; Table 3, aOR 0.08-0.70, *P*<.001). When adjusting for differences between settings, general measures were less likely to be adopted by male respondents (aOR 0.82; 95% CI 0.71-0.95) but more likely to be adopted by the unemployed (aOR 1.65; 95% CI 1.30-2.09) or retired (aOR 1.92; 95% CI 1.43-2.59). Contact measures were less likely to be adopted by male respondents (aOR 0.74; 95% CI:0.63-0.88) but more likely to be adopted by those who were retired (aOR 1.40; 95% CI 1.03-1.91). Finally, work measures were less likely to be adopted by respondents aged ≥55 years (aOR 0.60; 95% CI 0.39-0.93).

Table 3. Factors associated with the adoption of different social distancing measures.

Factor	Types of social distancing measures					
	General[a] (n=3431) (Model 1)		Contact[b] (n=3431) (Model 2)		Work[c] (n=2160) (Model 3)	
	aOR[d] (95% CI)	*P* value	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value
Age (years)						
18-24	Reference	N/A[e]	Reference	N/A	Reference	N/A
25-34	1.54 (1.16-2.05)	.003	0.96 (0.68-1.35)	.81	1.00 (0.72-1.39)	.99
35-44	1.25 (0.93-1.68)	.13	0.74 (0.53-1.05)	.09	0.95 (0.68-1.33)	.77
45-54	1.30 (0.95-1.79)	.10	0.67 (0.47-0.97)	.03	0.72 (0.49-1.05)	.09
55+	0.99 (0.70-1.41)	.97	0.81 (0.55-1.19)	.28	0.60 (0.39-0.93)	.02
Sex						
Female	Reference	N/A	Reference	N/A	Reference	N/A
Male	0.82 (0.71-0.95)	.01	0.74 (0.63-0.88)	<.001	0.95 (0.78-1.16)	.62
Education attainment						
No formal qualification/lower secondary or below	Reference	N/A	Reference	N/A	Reference	N/A
Secondary level qualification/higher secondary	0.99 (0.68-1.44)	.96	0.96 (0.64-1.44)	.85	1.04 (0.46-2.34)	.92
Postsecondary but below degree	1.06 (0.72-1.55)	.78	1.12 (0.74-1.71)	.58	1.09 (0.48-2.47)	.83
Degree or above	1.27 (0.87-1.83)	.21	0.98 (0.66-1.47)	.94	1.94 (0.88-4.28)	.10
Employment status						
Employed	Reference	N/A	Reference	N/A	N/A	N/A
Full-time student	1.35 (0.98-1.85)	.07	1.08 (0.73-1.59)	.70	N/A	N/A
Unemployed	1.65 (1.30-2.09)	<.001	1.20 (0.91-1.58)	.20	N/A	N/A
Retired	1.92 (1.43-2.59)	<.001	1.40 (1.03-1.91)	.03	N/A	N/A
Setting						
Hong Kong	Reference	N/A	Reference	N/A	Reference	N/A
United Kingdom	0.35 (0.30-0.41)	<.001	0.08 (0.07-0.10)	<.001	0.70 (0.57-0.87)	<.001

[a]General: avoiding going to crowded areas; going to social events; and going out.

[b]Contact: avoiding contacting individuals who had a fever or respiratory symptoms and had been to Wuhan in the past month (Hong Kong)/affected areas in the past 14 days (United Kingdom).

[c]Work: avoiding going to work.

[d]aOR: adjusted odds ratio.

[e]N/A: not applicable.

The impact of perceived severity of infection (Table 4) and perceived ease of transmission (Table 5) on the adoption of social distancing behaviors varied by setting. In Hong Kong, these factors had no impact, whereas in the United Kingdom, those who perceived COVID-19 infection as serious were more likely to adopt all social distancing measures (aOR 1.58-3.01), and those who perceived transmission of SARS-CoV-2 as easy were more likely to adopt both general social distancing measures (Model 4.2, aOR 2.00, 95% CI 1.57-2.55) and contact measures (Model 5.2, aOR 1.80, 95% CI 1.41-2.30). On the other hand, the impact of anxiety on the adoption of social distancing behaviors did not significantly differ by setting (Table 6). Those with borderline abnormal (Hong Kong, aOR 1.26-1.62; United Kingdom, aOR 1.36-1.48) or abnormal (Hong Kong, aOR:1.82-2.09; United Kingdom, aOR:1.40-2.40) HADS-A scores were more likely to adopt all three types of social distancing measures compared to those with normal anxiety levels.

Table 4. Setting-specific effects and effect modification (by setting) of perceived severity on the adoption of social distancing measures. The models have been adjusted for all covariates.

Settings and variables	Types of social distancing					
	Model 4.1		Model 5.1		Model 6.1	
	General[a] (n=3431)		Contact[b] (n=3431)		Work[c] (n=2160)	
	aOR[d] (95% CI)	*P* value	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value
Hong Kong						
Not serious[e]	Reference	N/A[f]	Reference	N/A	Reference	N/A
Serious[g]	0.93 (0.50-1.74)	.82	1.71 (0.85-3.47)	.13	0.63 (0.30-1.30)	.21
United Kingdom						
Not serious	Reference	N/A	Reference	N/A	Reference	N/A
Serious	3.01 (2.35-3.86)	<.001	1.90 (1.48-2.43)	<.001	1.58 (1.06-2.37)	.03
Effect modification[h]	3.24 (1.65-6.35)	<.001	1.11 (0.52-2.34)	.79	2.52 (1.09-5.80)	.03

[a]General: avoiding going to crowded areas and social events and going out.

[b]Contact: avoiding contacting individuals who had a fever or respiratory symptoms and had been to Wuhan in the past one month (Hong Kong) or affected areas in the past 14 days (United Kingdom).

[c]Work: avoiding going to work.

[d]aOR: adjusted odds ratio.

[e]For perceived severity, “not serious” refers to levels 3-5 (neutral to not serious at all, Hong Kong; moderate to no symptoms, United Kingdom).

[f]N/A: not applicable.

[g]For perceived severity, “serious” refers to levels 1-2 (very serious to serious, Hong Kong; life-threatening to severe, United Kingdom).

[h]Measures the difference of the effect being considered due to difference in setting; its value is the ratio of the two setting-specific effects.

Table 5. Setting-specific effects and effect modification (by setting) of perceived ease of transmission on the adoption of social distancing measures. The models have been adjusted for all covariates.

Settings and variables	Types of social distancing					
	Model 4.2		Model 5.2		Model 6.2	
	General[a] (n=3431)		Contact[b] (n=3431)		Work[c] (n=2160)	
	aOR[d] (95% CI)	*P* value	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value
Hong Kong						
Difficult[e]	Reference	N/A[f]	Reference	N/A	Reference	N/A
Easy[g]	1.15 (0.77-1.74)	.50	1.00 (0.58-1.71)	.99	0.61 (0.37-1.03)	.07
United Kingdom						
Difficult	Reference	N/A	Reference	N/A	Reference	N/A
Easy	2.00 (1.57-2.55)	<.001	1.80 (1.41-2.30)	<.001	1.34 (0.96-1.87)	.09
Effect modification[h]	1.73 (1.07-2.79)	.02	1.81 (1.00-3.28)	.05	2.18 (1.18- 4.04)	.01

[a]General: avoiding going to crowded areas and social events and going out.

[b]Contact: avoiding contacting individuals who had a fever or respiratory symptoms and had been to Wuhan in the past one month (Hong Kong) or affected areas in the past 14 days (United Kingdom).

[c]Work: avoiding going to work.

[d]aOR: adjusted odds ratio.

[e]For perceived ease of transmission, “difficult” means that the virus cannot be transmitted by face-to face conversation with someone who has SARS-CoV-2 but no symptoms.

[f]N/A: not applicable.

[g]For perceived ease of transmission, “easy” means that the virus can be transmitted by face-to face conversation with someone who has SARS-CoV-2 but no symptoms.

[h]Measures the difference of the effect being considered due to difference in setting; its value is the ratio of the two setting-specific effects.

Table 6. Setting-specific effects and effect modification (by setting) of anxiety level on the adoption of social distancing measures. The models have been adjusted for all covariates.

Settings and variables	Types of social distancing					
	Model 4.3		Model 5.3		Model 6.3	
	General[a] (n=3431)		Contact[b] (n=3431)		Work[c] (n=2160)	
	aOR[d] (95% CI)	*P* value	aOR (95% CI)	*P* value	aOR (95% CI)	*P* value
Hong Kong						
Normal	Reference	N/A[e]	Reference	N/A	Reference	N/A
Borderline abnormal	1.62 (1.27-2.06)	<.001	1.26 (0.93-1.70)	.14	1.51 (1.10-2.06)	.01
Abnormal	2.09 (1.64-2.66)	<.001	1.85 (1.34-2.56)	<.001	1.82 (1.34-2.48)	<.001
United Kingdom						
Normal	Reference	N/A	Reference	N/A	Reference	N/A
Borderline abnormal	1.48 (1.11-1.96)	.01	1.36 (1.02-1.80)	.03	1.37 (0.92-2.02)	.12
Abnormal	2.40 (1.87-3.09)	<.001	1.76 (1.37-2.27)	<.001	1.40 (0.99-1.98)	.06
Effect modification[f] (borderline abnormal)	0.91 (0.63-1.33)	.64	1.08 (0.71, 1.63)	.71	0.91 (0.55-1.49)	.70
Effect modification (abnormal)	1.15 (0.82-1.62)	.42	0.95 (0.63-1.42)	.80	0.77 (0.48-1.21)	.26

[a]General: avoiding going to crowded areas and social events and going out.

[b]Contact: avoiding contacting individuals who had a fever or respiratory symptoms and had been to Wuhan in the past one month (Hong Kong) or affected areas in the past 14 days (United Kingdom).

[c]Work: avoiding going to work.

[d]aOR: adjusted odds ratio.

[e]N/A: not applicable.

[f]Measures the difference of the effect being considered due to difference in setting; its value is the ratio of the two setting-specific effects.

Discussion

Principal Results

This study compared the initial public perceptions and preventative behaviors during the COVID-19 pandemic across Hong Kong and the United Kingdom. The adoption of social-distancing measures was higher in Hong Kong than in the United Kingdom. Risk perception and knowledge of COVID-19 were consistently and significantly higher in Hong Kong; however, for the United Kingdom, respondents' adoption of preventive behaviors was associated with two metrics: if transmission was considered to be "easy" and the perceived severity was "severe," UK respondents were more likely to adopt preventive behaviors. Anxiety was a driver of behavior change in both settings: those who were more anxious were more likely to adopt preventative measures. This behavior is consistent with the wider literature surrounding the adoption of precautionary measures [6,7] and provides further evidence that anxiety drives protective behaviors, such as handwashing [8], an effective intervention against the transmission of respiratory diseases [28].

Implications

This study has three implications. First, health officials should account for context-specific baseline levels of risk perception and knowledge when designing and promoting mitigation strategies. The evidence presented in this study demonstrates that geographical and sociocultural context is important in terms of both how people understand risk and how risk drives behavior. Although the social, historical, and cultural heterogeneity between Hong Kong and the United Kingdom likely contributes to the results of this study, the importance of intrinsic factors such as population sensitization via past infectious disease outbreaks and state-led health promotion campaigns should not be underestimated. In other studies, public perceptions of these factors have been found to be significant drivers of adopting preventative behaviors during previous epidemics [29], while conceptions of personal risk have been connected to individuals' understanding of local disease prevalence and severity [30-32]. Therefore, assessment of the baseline population knowledge, attitudes, and practices (KAP) and subsequent continual monitoring throughout the pandemic are essential to effective context-specific pandemic preparedness plans.

Second, risk communication should build upon baseline KAP outcomes, and trust should be developed across suitable media channels. Significant contextual heterogeneity in the public reliance on information sources provides insight here. Hong Kong reported greater reliance on social media and far less trust in official websites, suggesting that official messaging in Hong

Kong did not likely drive individual behavior change; by contrast, the UK results suggested that although the UK government possessed an effective platform to influence public health behavior, government health messaging was insufficient to attain similar baseline knowledge levels to those in Hong Kong, particularly in the absence of prior population sensitization to infectious disease outbreaks. Therefore, there is a pressing need to tailor communication approaches, likely on a graduated scale, but at a minimum in a binary fashion to accommodate both "naive" and "experienced" populations.

Third, the comparative snapshots of initial community responses captured by this study demonstrate the diversity in approaches and pandemic responses during the early phases of the COVID-19 pandemic. Across many contexts, national lockdowns became commonplace as the true magnitude of transmission became apparent; however, the associated indirect costs render blanket strategies untenable in the medium term. As national lockdowns are lifted, countries worldwide face the challenge of resurging cases and must consider nuanced approaches to prevent additional harm. Driven by anxiety, high perceived severity and knowledge, Hong Kong conducted widespread preventive measures early and en masse. Together with early government actions [33], the strategies adopted by the Hong Kong community were successful during the initial phase of the pandemic. Considering this—and that national populations are now highly sensitized to COVID-19 transmission—tailored public health messaging, early regional containment, and increased health capacity should ensure more effective public health responses with less indirect impact on national economies.

Study Strengths and Limitations

From a methodological perspective, the UK sampling approach enabled the sample size to be achieved quickly, thereby accurately capturing prevailing sentiment and behavior across a short time frame (2 days). However, this approach likely came at the expense of excluding participants without access to the internet, and it contrasted with the survey period in Hong Kong (3 weeks); this likely led to some sampling bias, especially during the initial phase of the pandemic (when there was much uncertainty about the disease). Additionally, both samples varied across the demographic spectrum; thus, although responses were weighted, caution should be taken when extrapolating study findings to wider populations. Moreover, given the incompatibility of region-specific weights and the controversy in estimating standard errors when survey weights are involved [34], unweighted regression results were presented; however, they should be interpreted with caution. Last but not least, although both surveys were conducted early locally, the difference in surrounding international events during the survey periods (eg, the Hong Kong survey was conducted before COVID-19 was formally declared a pandemic, but the UK survey was launched after this declaration) may have introduced bias in the survey responses.

Conclusions

This study compared the initial community responses to COVID-19 in Hong Kong and the United Kingdom. In line with the high baseline level of risk perception and knowledge and with historical exposure to respiratory disease outbreaks, the adoption of preventive measures was higher in Hong Kong. However, the UK sample demonstrated that this adoption could be improved by heightened risk perception and knowledge, best driven by improved public health campaigns. Together, these results suggest that health officials should ascertain baseline levels of risk perception and knowledge, as well as prior sensitization to infectious disease outbreaks, when developing mitigation strategies. Risk communication should be performed through suitable media channels—and trust should be maintained—while early intervention remains the cornerstone of effective outbreak response.

Acknowledgments

The study was supported by Imperial NIHR Research Capability Funding and the internal funding of The Chinese University of Hong Kong. HW is a National Institute for Health Research (NIHR) Senior Investigator and receives funding from the Imperial NIHR Biomedical Research Centre, the NIHR Applied Research Collaborative North West London, the NIHR School for Public Health Research, and the Wellcome Trust. KOK would like to acknowledge the Research Fund for the Control of Infectious Diseases, Hong Kong (number:INF-CUHK-1); General Research Fund (number:14112818); Early Career Scheme (number: 24104920); Health and Medical Research Fund (numbers:17160302, 18170312); The Wellcome Trust (number: 200861/Z/16/Z); and a Direct Grant for Research of The Chinese University of Hong Kong (CUHK) (number: 2019.020).

Authors' Contributions

LRB, KOK, HW, CA, and SYSW conceived the study; KOK, WIW, and CA collected the data; KOK, YYY, and WIW analyzed the data; LRB, KOK, RER, HW, WIW, CA, and SYSW interpreted the data; LRB wrote the first draft of the manuscript; and KOK, RER, YYY, HW, WIW, CA, and SYSW edited the manuscript. LRB and KOK contributed equally as the joint first author. CA and SYSW also contributed equally as the joint last author. KOK and CA also contributed equally as the joint corresponding author.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary files.
[DOCX File , 45 KB - jmir_v23i3e23231_app1.docx]

References

1. Rolling updates on coronavirus disease (COVID-19). World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen [accessed 2021-02-12]
2. Countries/areas with reported cases of Coronavirus Disease 2019 (COVID-19) (Last updated on August 2, 2020, 11 am) 2020. Hong Kong Centre for Health Protection. URL: https://www.chp.gov.hk/files/pdf/statistics_of_the_cases_novel_coronavirus_infection_en.pdf [accessed 2021-02-12]
3. Kwok KO, Lai FYL, Wei VWI, Tsoi MTF, Wong SYS, Tang JWT. Comparing the impact of various interventions to control the spread of COVID-19 in twelve countries. J Hosp Infect 2020 Sep;106(1):214-216 [FREE Full text] [doi: 10.1016/j.jhin.2020.06.029] [Medline: 32619456]
4. Prime Minister's statement on coronavirus (COVID-19): 16 March 2020. Prime Minister's Office of the Government of the United Kingdom. 2020 Mar 16. URL: https://www.gov.uk/government/speeches/pm-statement-on-coronavirus-16-march-2020 [accessed 2021-02-12]
5. Kwok KO, Li KK, Chan HHH, Yi YY, Tang A, Wei WI, et al. Community responses during early phase of COVID-19 epidemic, Hong Kong. Emerg Infect Dis 2020 Jul;26(7):1575-1579 [FREE Full text] [doi: 10.3201/eid2607.200500] [Medline: 32298227]
6. Leung GM, Ho L, Chan SKK, Ho S, Bacon-Shone J, Choy RYL, et al. Longitudinal assessment of community psychobehavioral responses during and after the 2003 outbreak of severe acute respiratory syndrome in Hong Kong. Clin Infect Dis 2005 Jun 15;40(12):1713-1720. [doi: 10.1086/429923] [Medline: 15909256]
7. Bish A, Michie S. Demographic and attitudinal determinants of protective behaviours during a pandemic: a review. Br J Health Psychol 2010 Nov;15(Pt 4):797-824 [FREE Full text] [doi: 10.1348/135910710X485826] [Medline: 20109274]
8. Jones JH, Salathé M. Early assessment of anxiety and behavioral response to novel swine-origin influenza A (H1N1). PLoS One 2009;4(12):e8032 [FREE Full text] [doi: 10.1371/journal.pone.0008032] [Medline: 19997505]
9. Bults M, Beaujean DJ, de Zwart O, Kok G, van Empelen P, van Steenbergen JE, et al. Perceived risk, anxiety, and behavioural responses of the general public during the early phase of the Influenza A (H1N1) pandemic in the Netherlands: results of three consecutive online surveys. BMC Public Health 2011 Jan 03;11:2 [FREE Full text] [doi: 10.1186/1471-2458-11-2] [Medline: 21199571]
10. Lohiniva A, Sane J, Sibenberg K, Puumalainen T, Salminen M. Understanding coronavirus disease (COVID-19) risk perceptions among the public to enhance risk communication efforts: a practical approach for outbreaks, Finland, February 2020. Euro Surveill 2020 Apr;25(13):2000317 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.13.2000317] [Medline: 32265008]
11. Gesser-Edelsburg A, Cohen R, Hijazi R, Abed Elhadi Shahbari N. Analysis of public perception of the Israeli government's early emergency instructions regarding COVID-19: online survey study. J Med Internet Res 2020 May 15;22(5):e19370 [FREE Full text] [doi: 10.2196/19370] [Medline: 32392172]
12. Motta Zanin G, Gentile E, Parisi A, Spasiano D. A preliminary evaluation of the public risk perception related to the COVID-19 health emergency in Italy. Int J Environ Res Public Health 2020 Apr 27;17(9):3024 [FREE Full text] [doi: 10.3390/ijerph17093024] [Medline: 32349253]
13. Olapegba PO, Iorfa SK, Kolawole SQ, Oguntayo R, Gandi JC, Ottu IFA, et al. Survey data of COVID-19-related knowledge, risk perceptions and precautionary behavior among Nigerians. Data Brief 2020 Jun;30:105685 [FREE Full text] [doi: 10.1016/j.dib.2020.105685] [Medline: 32391411]
14. McFadden SM, Malik AA, Aguolu OG, Willebrand KS, Omer SB. Perceptions of the adult US population regarding the novel coronavirus outbreak. PLoS One 2020;15(4):e0231808 [FREE Full text] [doi: 10.1371/journal.pone.0231808] [Medline: 32302370]
15. Clements JM. Knowledge and behaviors toward COVID-19 among US residents during the early days of the pandemic: cross-sectional online questionnaire. JMIR Public Health Surveill 2020 May 08;6(2):e19161 [FREE Full text] [doi: 10.2196/19161] [Medline: 32369759]
16. Husnayain A, Shim E, Fuad A, Su EC. Understanding the community risk perceptions of the COVID-19 outbreak in South Korea: infodemiology study. J Med Internet Res 2020 Sep 29;22(9):e19788 [FREE Full text] [doi: 10.2196/19788] [Medline: 32931446]
17. Huynh TLD. Data for understanding the risk perception of COVID-19 from Vietnamese sample. Data Brief 2020 Jun;30:105530 [FREE Full text] [doi: 10.1016/j.dib.2020.105530] [Medline: 32322641]
18. Christina JA, Leigh B, Charlotte V, Rozlyn R, Philippa P, Jeffrey WE, et al. Perceptions and behavioural responses of the general public during the COVID-19 pandemic: A cross-sectional survey of UK Adults. medRxiv (forthcoming) [FREE Full text] [doi: 10.1101/2020.04.01.20050039]
19. Latest situation of cases of COVID-19. Hong Kong Centre for Health Protection. URL: https://www.chp.gov.hk/files/pdf/local_situation_covid19_en.pdf [accessed 2020-08-02]

20. Coronavirus (COVID-19) in the UK: Cases in United Kingdom. Government of the United Kingdom. URL: https://coronavirus.data.gov.uk/cases [accessed 2020-10-19]
21. ESOMAR 28: 28 questions to help online research buyers. YouGov. URL: http://doc.ukdataservice.ac.uk/doc/7647/mrdoc/pdf/7647_methodology_esomar_28.pdf [accessed 2020-10-18]
22. COVID 19 - population survey tool. The Chinese University of Hong Kong. URL: https://www.sphpc.cuhk.edu.hk/covid19-populationsurveytools [accessed 2020-10-18]
23. COVID-19 scientific resources. MRC Centre for Global Infectious Disease Analysis, Imperial College London. URL: http://www.imperial.ac.uk/medicine/departments/school-public-health/infectious-disease-epidemiology/mrc-global-infectious-disease-analysis/covid-19/covid-19-scientific-resources/ [accessed 2020-10-18]
24. Snaith RP. The Hospital Anxiety And Depression Scale. Health Qual Life Outcomes 2003 Aug 01;1:29 [FREE Full text] [doi: 10.1186/1477-7525-1-29] [Medline: 12914662]
25. 2011 Census. UK Office for National Statistics. URL: https://www.ons.gov.uk/census/2011census [accessed 2021-02-12]
26. By-census results 2016. Hong Kong Census and Statistics Department. URL: https://www.bycensus2016.gov.hk/en/ [accessed 2021-02-12]
27. Anderson L, Fricker RD. Raking: an important and often overlooked survey analysis tool. Phalanx 2015;48(3):36-42.
28. Rabie T, Curtis V. Handwashing and risk of respiratory infections: a quantitative systematic review. Trop Med Int Health 2006 Mar;11(3):258-267. [doi: 10.1111/j.1365-3156.2006.01568.x] [Medline: 16553905]
29. SteelFisher GK, Blendon RJ, Ward JRM, Rapoport R, Kahn EB, Kohl KS. Public response to the 2009 influenza A H1N1 pandemic: a polling study in five countries. Lancet Infect Dis 2012 Nov;12(11):845-850. [doi: 10.1016/S1473-3099(12)70206-2] [Medline: 23041199]
30. Funk S, Salathé M, Jansen VAA. Modelling the influence of human behaviour on the spread of infectious diseases: a review. J R Soc Interface 2010 Sep 06;7(50):1247-1256 [FREE Full text] [doi: 10.1098/rsif.2010.0142] [Medline: 20504800]
31. de Zwart O, Veldhuijzen IK, Elam G, Aro AR, Abraham T, Bishop GD, et al. Perceived threat, risk perception, and efficacy beliefs related to SARS and other (emerging) infectious diseases: results of an international survey. Int J Behav Med 2009;16(1):30-40 [FREE Full text] [doi: 10.1007/s12529-008-9008-2] [Medline: 19125335]
32. Bults M, Beaujean DJMA, Richardus JH, Voeten HACM. Perceptions and behavioral responses of the general public during the 2009 influenza A (H1N1) pandemic: a systematic review. Disaster Med Public Health Prep 2015 Apr;9(2):207-219. [doi: 10.1017/dmp.2014.160] [Medline: 25882127]
33. Wong SYS, Kwok KO, Chan FKL. What can countries learn from Hong Kong's response to the COVID-19 pandemic? CMAJ 2020 May 11;192(19):E511-E515 [FREE Full text] [doi: 10.1503/cmaj.200563] [Medline: 32332040]
34. Gelman A. Struggles with survey weighting and regression modeling. Statist Sci 2007 May;22(2):153-164. [doi: 10.1214/088342306000000691]

Abbreviations

aOR: adjusted odds ratio
HASD-A: Hospital Anxiety and Depression scale–Anxiety
KAP: knowledge, attitudes, and practices

Edited by G Fagherazzi; submitted 05.08.20; peer-reviewed by W Camelo Castillo, B Prainsack; comments to author 19.09.20; revised version received 19.10.20; accepted 01.02.21; published 08.03.21.

Please cite as:
Bowman L, Kwok KO, Redd R, Yi Y, Ward H, Wei WI, Atchison C, Wong SYS
Comparing Public Perceptions and Preventive Behaviors During the Early Phase of the COVID-19 Pandemic in Hong Kong and the United Kingdom: Cross-sectional Survey Study
J Med Internet Res 2021;23(3):e23231
URL: https://www.jmir.org/2021/3/e23231
doi:10.2196/23231
PMID:33539309

Original Paper

A Comprehensive Overview of the COVID-19 Literature: Machine Learning–Based Bibliometric Analysis

Alaa Abd-Alrazaq[1], PhD; Jens Schneider[1], PhD; Borbala Mifsud[2], PhD; Tanvir Alam[1], PhD; Mowafa Househ[1], PhD; Mounir Hamdi[1], PhD; Zubair Shah[1], PhD

[1]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

[2]College of Health and Life Sciences, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

Corresponding Author:
Zubair Shah, PhD
Division of Information and Computing Technology, College of Science and Engineering
Hamad Bin Khalifa University
Qatar Foundation
P.O. Box 5825, Doha Al Luqta St, Ar-Rayyan
Doha, 00000
Qatar
Phone: 974 55708549
Email: zshah@hbku.edu.qa

Abstract

Background: Shortly after the emergence of COVID-19, researchers rapidly mobilized to study numerous aspects of the disease such as its evolution, clinical manifestations, effects, treatments, and vaccinations. This led to a rapid increase in the number of COVID-19–related publications. Identifying trends and areas of interest using traditional review methods (eg, scoping and systematic reviews) for such a large domain area is challenging.

Objective: We aimed to conduct an extensive bibliometric analysis to provide a comprehensive overview of the COVID-19 literature.

Methods: We used the COVID-19 Open Research Dataset (CORD-19) that consists of a large number of research articles related to all coronaviruses. We used a machine learning–based method to analyze the most relevant COVID-19–related articles and extracted the most prominent topics. Specifically, we used a clustering algorithm to group published articles based on the similarity of their abstracts to identify research hotspots and current research directions. We have made our software accessible to the community via GitHub.

Results: Of the 196,630 publications retrieved from the database, we included 28,904 in our analysis. The mean number of weekly publications was 990 (SD 789.3). The country that published the highest number of COVID-19–related articles was China (2950/17,270, 17.08%). The highest number of articles were published in bioRxiv. Lei Liu affiliated with the Southern University of Science and Technology in China published the highest number of articles (n=46). Based on titles and abstracts alone, we were able to identify 1515 surveys, 733 systematic reviews, 512 cohort studies, 480 meta-analyses, and 362 randomized control trials. We identified 19 different topics covered among the publications reviewed. The most dominant topic was public health response, followed by clinical care practices during the COVID-19 pandemic, clinical characteristics and risk factors, and epidemic models for its spread.

Conclusions: We provide an overview of the COVID-19 literature and have identified current hotspots and research directions. Our findings can be useful for the research community to help prioritize research needs and recognize leading COVID-19 researchers, institutes, countries, and publishers. Our study shows that an AI-based bibliometric analysis has the potential to rapidly explore a large corpus of academic publications during a public health crisis. We believe that this work can be used to analyze other eHealth-related literature to help clinicians, administrators, and policy makers to obtain a holistic view of the literature and be able to categorize different topics of the existing research for further analyses. It can be further scaled (for instance, in time) to clinical summary documentation. Publishers should avoid noise in the data by developing a way to trace the evolution of individual publications and unique authors.

(J Med Internet Res 2021;23(3):e23703) doi:10.2196/23703

KEYWORDS

novel coronavirus disease; COVID-19; SARS-CoV-2; 2019-nCoV; bibliometric analysis; literature; machine learning; research; review

Introduction

Background

In December 2019, Wuhan city in China registered several cases of an unknown disease characterized by pneumonia, dry cough, fatigue, and fever [1]. The investigations revealed that a novel coronavirus (2019-nCoV) was the causative agent of the disease, which was subsequently named COVID-19 [1]. Since then, COVID-19 has spread around the globe, leading the World Health Organization to classify it as a pandemic [2]. This highly contagious pathogen has affected almost every aspect of our daily lives, such as education, traveling, business, transportation, sports, and health care [3]. Most importantly, the COVID-19 pandemic has claimed more than 775,000 lives as of August 19, 2020 [4]. To curb the impact of COVID-19, authorities need to implement effective public health measures related to COVID-19 surveillance, diagnostics, vaccines, treatments, and research [5].

Given the novelty and, consequently, the lack of knowledge about the disease, research can play a crucial role in the fight against the COVID-19 pandemic. Scientists have rapidly mobilized to manage and slowdown the growth of the pandemic. The scientific literature in this domain area has exponentially increased [6,7]. By the end of May 2020, Aristovnik et al [6] and Doanvo et al [8] retrieved 10,344 and 18,412 COVID-19–related publications written in the English language, respectively, from the Scopus database and the COVID-19 Open Research Dataset (CORD-19). In addition, as of July 13, 2020, more than 1711 clinical trials were registered in different clinical trial registries (eg, NCT, EUCTR, and ISRCTN) [9].

It is very important to have a comprehensive overview of the current state of the literature on COVID-19 for several reasons, namely: (1) to organize and coordinate the literature; (2) to explore research topics addressed; (3) to prioritize research needs or gaps; (4) to understand the evolution of the literature; (5) to recognize the leading researchers, institutes, and countries in this area; and (6) to explore connections between research topics and areas.

Research Problem and Aim

Manually conducting a comprehensive review of the thousands of COVID-19–related publications is a daunting and time-consuming task. Artificial intelligence (AI) methods can play a pivotal role in rapidly surveying the enormous number of publications and extracting critical insights from them. Therefore, in March 2020, the White House strongly recommended researchers to exploit AI methods in COVID-19 research [8].

Several studies have used AI methods to conduct a bibliometric analysis of research on the COVID-19 pandemic [6-8,10-12]. However, we identified the following research gaps in these studies. First, publications analyzed in most studies were dated, approximately to the first three months after the onset of the COVID-19 outbreak; thus, numerous studies published afterward were not analyzed [7,10-16]. Second, several studies analyzed publications related to all types of coronaviruses instead of focusing on COVID-19 [7,10-12,17,18]; hence, the results related to COVID-19 were aggregated with those related to other coronaviruses. Third, several studies included only a few (ranging from 38 to 1482) of the large number of publications related to COVID-19 available in the search period [7,10,12-16,19]. Fourth, most studies did not examine the topics that previous studied had addressed, instead they assessed only the metadata of those studies (eg, countries, authors, number of citations, and published journals) [13,14,16-19]. Fifth, topic identification among various studies was conducted using manual screening instead of AI methods [13-15]. To fill the abovementioned gaps, this study aims to conduct an extensive bibliometric analysis to provide a comprehensive overview of the existing COVID-19 literature.

Methods

Study Data Collection

For this study, we used CORD-19, generated by the Allen Institute for AI [20]. The dataset is updated daily to include the latest published articles on COVID-19. We used the update corresponding to the timestamp of July 21, 2020, which contained over 196,630 scholarly articles related to COVID-19 and the coronavirus family of viruses. Allen Institute for AI used the following search terms to retrieve studies on all coronaviruses: “COVID-19” OR “Coronavirus” OR “Corona virus” OR “2019-nCoV” OR “SARS-CoV” OR “MERS-CoV” OR “Severe Acute Respiratory Syndrome” OR “Middle East Respiratory Syndrome”. The search was conducted on PubMed, PubMed Central, and bioRxiv and medRxiv preprint servers. The dataset included a CSV (comma-separate values) file with metadata of all the articles in the dataset, such as article ID, title, abstract, names of authors, and publication date. The articles in the dataset were represented by a single JSON (JavaScript Object Notation) file that consisted of the article ID, title, abstract, body text, and relevant metadata. The metadata of the dataset was analyzed using Python in a Jupyter Notebook environment. We have made our software accessible to the community via GitHub [21]. The CSV metadata file was loaded into a data frame provided by Python’s pandas library. We removed records with empty and non-English abstracts. We also removed duplicate articles and any articles that were published before January 1, 2020. We then used the search terms “novel coronavirus,” “coronavirus 2019,” “2019-nCov,” “COVID-19,” “COVID 2019,” “severe acute respiratory syndrome coronavirus 2,” and “SARS-COV-2” to select only COVID-19–related articles. Thus, we were able to identify a total of 28,904 abstracts of scholarly articles published after January 1, 2020, that were related to COVID-19 for the downstream analysis.

Data Preprocessing

The 28,904 selected abstracts were cleaned by removing punctuations and alphanumeric characters. Singular and plural uppercased abstract sectioning keywords such as "BACKGROUND," "OBJECTIVE," "METHOD," "RESULT," and "CONCLUSION" were also removed. The data cleaning was performed using Python programming language in Jupyter Notebook environment. The Python libraries used to clean the data include pandas, NumPy, langdetect, re, string, and TextBlob. The abstracts were then converted to lowercased text. After that, we used the Python Natural Language Toolkit library to tokenize the abstracts and remove the stop words. We then applied the SnowballStemmer model to convert words to their stems. The clean text of the abstracts derived after applying the abovementioned pre-processing steps was used for clustering.

Document Clustering

For document clustering, we first converted each document (ie, abstract) to a feature vector, where features were defined by term (ie, words) frequency–inverse document frequency (TF-IDF) weights. TF-IDF represents the importance of a word relative to a document in a corpus. This importance increases proportionally to the number of times the word appears in the document but is offset by the frequency of that word in the corpus. This ensures that TF-IDF–based similarity measures between documents are influenced mainly by discriminative words with relatively low frequencies in the corpus [22]. For TF-IDF representation of the abstracts, we used TfidfVectorizer module of the Python scikit library.

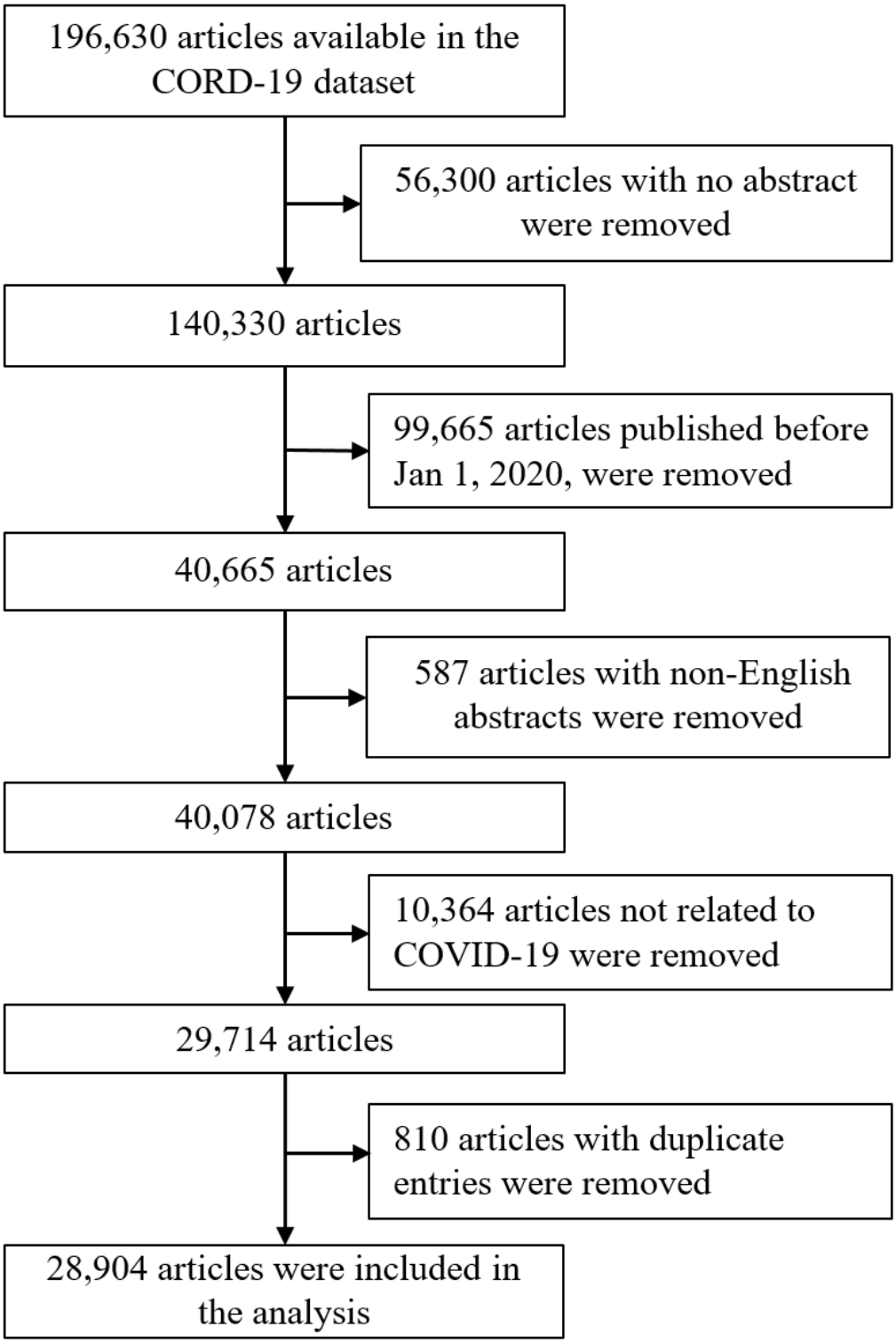
The TfidfVectorizer algorithm has two important threshold parameters that cut off low and high word frequencies. The minimum document frequency parameter (min_df) was set to 10 to ignore sporadic terms occurring in less than 10 documents (absolute count). The maximum document frequency parameter (max_df) was set to 0.9 to ignore terms that appear in more than 90% (26,014/28,904) of the documents (relative count). The reason is that we wished to exclude terms that are either too rare to be used in finding document clusters or too common to be discriminative enough to distinguish documents. Based on these parameters (min_df and max_df), the TfidfVectorizer algorithm extracted a vector of 42,061 unique terms to represent each of the 28,904 abstracts, with each term containing a TF-IDF score. This generated a feature matrix of size 28,904×42,061, which was, subsequently, used to feed into a clustering algorithm. We used the k-means clustering algorithm from Python's scikit library to categorize the abstracts into internally coherent but well-separated clusters. To identify the number of clusters in the k-means clustering algorithm, we used the elbow method to determine the number of clusters in the corpus [23]. Thus, we found 26 to be the optimal number of clusters for this corpus.

Results

Search Results

By July 21, 2020, the CORD-19 dataset comprised 196,630 articles (Figure 1). Of those, we excluded 167,726 articles for the following reasons: (1) abstracts were unavailable (n=56,300); (2) the articles were published before January 1, 2020 (n=99,665); (3) the articles were written in a language other than English (n=587); (4) the articles were not related to COVID-19, as their titles and abstracts did not contain our search terms (n=10,364); and (5) they had duplicate entries (n=810). Consequently, we included 28,904 articles in the analysis in this study.

Figure 1. Flowchart of the selection of published articles. CORD-19: COVID-19 Open Research Dataset.



Characteristics of Publications

The first paper was published on January 2, 2020. As shown in Figure 2, the number of publications in each week increased considerably since then, until a peak was reached in week 22 (2276 publications). Thereafter, the number of research papers published began to decrease. The mean number of publications for each week was 990 (SD 789.3). The country of publication was identified for 17,270 publications, which were conducted across 221 countries and territories. The country that published the highest number of articles was China (2950/17,270, 17.08%), followed by the United States (1357/17,270, 7.86%), Italy (1157/17,270, 6.70%), Saudi Arabia (978/17,270, 5.66%), and India (854/17,270, 4.94%) (Table 1).

The selected articles were published in about 2500 journals. The highest number of articles were published in bioRxiv (n=1374), the most prominent preprint server for biology. The top 10 sites for publishing COVID-19–related articles (journals

and preprint servers) are shown in Table 2. The publications included in this analysis were authored by 150,600 authors. Among those authors, Lei Liu published the highest number of articles (n=46; see Table 3). Based on titles and abstracts alone, we were able to identify 1515 surveys, 733 systematic reviews, 512 cohort studies, 480 meta-analyses, 362 randomized control trials, 199 case studies, 79 scoping reviews, and 62 case-control studies (Table 4). Note that these numbers include only the top 8 study methods for those publications that mention the study method in either the abstract or the title.

Figure 2. Number of publications in each week (2020).

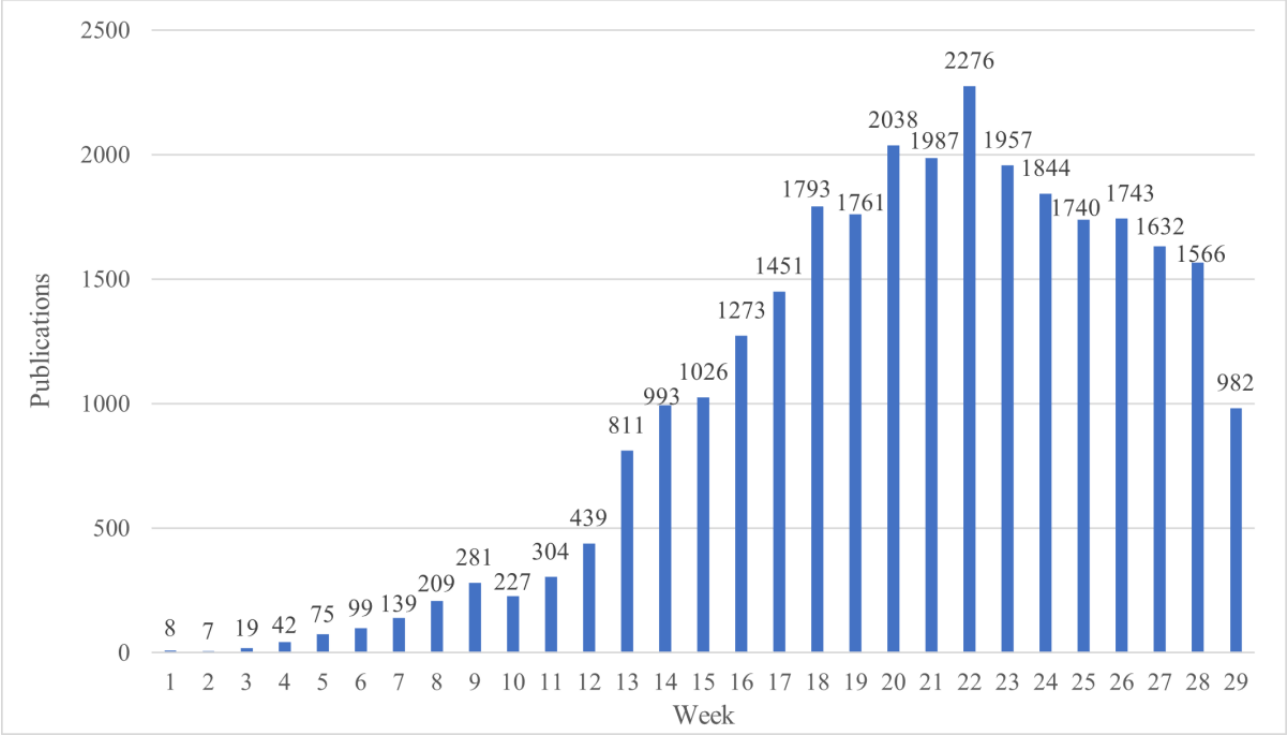


Table 1. Top 10 countries (N=221) by the number of COVID-19–related publications (N=17,270).

Rank	Country	Publications, n (%)
1	China	2950 (17.08)
2	United States	1357 (7.86)
3	Italy	1157 (6.70)
4	Saudi Arabia	978 (5.66)
5	India	854 (4.94)
6	Canada	671 (3.89)
7	United Kingdom	525 (3.04)
8	Germany	449 (2.60)
9	Australia	403 (2.33)
10	France	383 (2.22)

Table 2. Top 10 sites (journals and preprint servers) where COVID-19–related articles (N=28,904) were published.

Rank	Journal or server	Publications, n (%)
1	bioRxiv	1374 (4.75)
2	Journal of Medical Virology	468 (1.6)
3	medRxiv	340 (1.18)
4	International Journal of Environmental Research and Public Health	223 (0.77)
5	Clinical Infectious Diseases	195 (0.67)
6	International Journal of Infectious Diseases	184 (0.64)
7	Science of the Total Environment	165 (0.57)
8	Cureus	148 (0.51)
9	Psychological Trauma: Theory, Research, Practice, and Policy	138 (0.48)
10	Medical Hypotheses	136 (0.47)

Table 3. Top 10 authors (N=15,600) by the number of COVID-19–related publications.

Rank	Author name	Publications, n	Author affiliation
1	Lei Liu	46	Southern University of Science and Technology, China
2	Kwok-Yung Yuen	34	The University of Hong Kong, China
3	Christian Drosten	31	Charité University Medicine Berlin, Germany
4	Ralph S Baric	31	University of North Carolina at Chapel Hill, USA
5	Gerardo Chowell	30	Georgia State University, USA
6	Hongzhou Lu	30	Fudan University, Shanghai, China
7	Giuseppe Lippi	29	University of Verona, Italy
8	Jasper Fuk-Wo Chan	27	The University of Hong Kong, China
9	Kelvin Kai-Wang To	25	The University of Hong Kong, China
10	Valerie A Canady	23	Mental Health Weekly, USA

Table 4. Eight most common study methods extracted from the publications mentioning the study design used in the abstract.

Rank	Study method	Publications, n
1	Survey	1515
2	Systematic review	733
3	Cohort study	512
4	Meta-analysis	480
5	Randomized control trial	362
6	Case study	199
7	Scoping review	79
8	Case-control study	62

Results of Topics Modelling

Overview

The analysis generated 26 clusters from the included publications. We were able to identify the topic of 21 clusters, whereas the remaining 5 clusters were not labeled as they contained publications with very diverse topics that belonged to other clusters. Therefore, publications in these 5 clusters were moved to the most appropriate cluster among the 21 clusters. Four of the 21 clusters contained publications addressing only two different topics; thus, we further merged the 4 clusters to form 2 different clusters. Overall, we identified 19 different topics addressed in the included publications (Table 5).

Table 5. COVID-19–related topics addressed by the included publications (N=28,904).

Number	Topic	Articles, n (%)	Top 20 unigrams
1	Public health response	5393 (18.66)	covid, pandemic, health, coronavirus, disease, public, data, world, spread, social, study, global, outbreak, countries, cases, measures, virus, sars, results, people
2	Clinical care practices for patients during the COVID-19 pandemic	5118 (17.71)	covid, pandemic, patients, care, disease, coronavirus, health, patient, infection, risk, healthcare, clinical, sars, cov, management, medical, respiratory, hospital, cases, acute
3	Clinical characteristics and risk factors of COVID-19	3313 (11.46)	covid, patients, disease, coronavirus, severe, clinical, study, results, age, data, risk, respiratory, hospital, sars, infection, cov, cases, acute, mortality, higher
4	Epidemic models for COVID-19 spread	2964 (10.25)	covid, model, data, cases, number, epidemic, disease, time, results, spread, pandemic, coronavirus, based, countries, infected, infection, rate, study, measures, outbreak
5	Therapies and vaccines for COVID-19	1845 (6.38)	sars, cov, covid, drug, coronavirus, drugs, viral, antiviral, virus, potential, disease, pandemic, respiratory, treatment, protease, molecular, compounds, severe, based, inhibitors
6	Host immune response	1837 (6.36)	covid, cov, sars, disease, coronavirus, infection, severe, respiratory, patients, acute, syndrome, immune, viral, cells, virus, clinical, inflammatory, response, pandemic, cell, cytokine
7	Diagnosis of COVID-19 using PCR[a]	1602 (5.54)	cov, sars, covid, pcr, positive, coronavirus, testing, respiratory, results, patients, rt, disease, infection, detection, samples, test, clinical, viral, time, negative
8	Mental health and disorders during the COVID-19 pandemic	915 (3.17)	covid, health, mental, pandemic, anxiety, psychological, study, coronavirus, stress, depression, results, social, disease, risk, impact, related, people, survey, symptoms, outbreak
9	Diagnosis of COVID-19 based on chest imaging	874 (3.02)	covid, ct, patients, chest, disease, coronavirus, pneumonia, clinical, diagnosis, results, imaging, findings, lung, cases, ground, glass, tomography, computed, features, images
10	Social distancing measures	868 (3)	covid, social, distancing, pandemic, measures, spread, contact, disease, data, health, model, number, transmission, cases, control, population, coronavirus, time, tracing, public
11	Virus genomics	816 (2.82)	sars, cov, coronavirus, virus, genome, covid, viral, analysis, sequences, respiratory, severe, pandemic, human, disease, sequence, acute, syndrome, genomes, china, study
12	Protein structures of 2019-nCoV[b]	706 (2.44)	cov, sars, spike, protein, binding, coronavirus, covid, receptor, virus, human, viral, pandemic, ace, domain, cell, vaccine, infection, disease, host, development
13	Host cell entry	584 (2.02)	ace, sars, cov, covid, angiotensin, enzyme, converting, receptor, coronavirus, disease, infection, expression, respiratory, cells, severe, human, syndrome, entry, cell, virus
14	Clinical care practices for patients with cancer	441 (1.53)	cancer, covid, patients, pandemic, treatment, risk, care, disease, coronavirus, infection, health, clinical, patient, management, cov, sars, severe, therapy, high, oncology
15	Detection of 2019-nCoV antibodies	411 (1.42)	cov, sars, igg, covid, antibodies, patients, antibody, igm, infection, results, positive, disease, coronavirus, samples, study, test, assay, serological, sensitivity, clinical
16	Personal protective equipment	350 (1.21)	covid, masks, mask, pandemic, protective, equipment, face, personal, respirators, health, coronavirus, healthcare, workers, sars, cov, respiratory, surgical, disease, study, ppe
17	Diabetes mellitus and COVID-19	336 (1.16)	covid, pandemic, patients, diabetes, disease, care, risk, coronavirus, health, nursing, nurses, infection, management, severe, results, data, study, clinical, high, methods
18	Pregnancy and childbirth during the COVID-19 pandemic	312 (1.08)	women, pregnant, covid, pregnancy, infection, sars, coronavirus, cov, disease, severe, clinical, maternal, patients, cases, data, transmission, respiratory, pandemic, delivery, results
19	Organ transplantation during the COVID-19 pandemic	219 (0.76)	covid, transplant, recipients, patients, disease, coronavirus, pandemic, cov, sars, infection, transplantation, kidney, severe, organ, clinical, risk, respiratory, acute, liver, patient

[a]PCR: polymerase chain reaction.

[b]2019-nCoV: novel coronavirus.

Topic 1: Public Health Response

This topic was addressed by 18.66% (5393/28,904) of the publications. The publications in this cluster mainly discussed how public health authorities in various countries responded to the COVID-19 pandemic (eg, [24-28]). The top 5 authors in terms of the highest number of publications related to this topic were Claudine McCarthy (n=8), Valerie A Canady (n=6), Alison Knopf (n=6), Alimuddin Zumla (n=6), and Nima Rezaei (n=6). The top 5 journals and preprint servers hosting the highest number of publishing articles related to this topic were the International Journal of Environmental Research and Public

Health (n=82), Science of the Total Environment (n=80), New Scientist (n=56), Journal of Medical Virology (n=53), and bioRxiv (n=50). The first paper related to this topic was published on January 10, 2020. The number of publications in each week increased significantly until it reached a peak in week 23 (n=434); it then decreased noticeably (Multimedia Appendix 1). The mean number of weekly publications in this cluster was 183.6 (SD 151.5).

Topic 2: Clinical Care Practices During the COVID-19 Pandemic

A total of 17.71% (5118/28,904) of all included publications were mainly about clinical care practices for non–COVID-19 patients during the COVID-19 pandemic (eg, [29-33]). The following authors published the highest number of publications related to this topic: Karthik Rajasekaran (n=14), Francesco Esperto (n=12), Raju Vaishya (n=9), Namrata Sharma (n=8), and Santosh G Honavar (n=8). The top 5 journals publishing articles related to this topic were Otolaryngology-Head and Neck Surgery (n=115), the Journal of the European Academy of Dermatology and Venereology (n=45), Cureus Journal of Medical Science (n=41), Anaesthesia (n=40), and World Neurosurgery (n=35). In this cluster, the first article was published on January 3, 2020. There was a considerable rise in the number of weekly publications from week 12 until it reached a peak in week 23 (n=479); this was followed by a sharp decrease (Multimedia Appendix 1). The mean number of weekly publications in this cluster was 175.2 (SD 159.6).

Topic 3: Clinical Characteristics and Risk Factors of COVID-19

This topic was discussed in 11.46% (3313/28,904) of the publications (eg, [34-38]). The top 5 authors who published the highest number of publications in this cluster were Lei Liu (n=25), Lanjuan Li (n=13), Jifang Sheng (n=10), Giuseppe Lippi (n=9), and Nanshan Zhong (n=9). The top 5 journals and preprint servers publishing articles related to this topic were the Journal of Medical Virology (n=109), medRxiv (n=57), Cureus Journal of Medical Science (n=35), Clinical Infectious Diseases (n=33), and the International Journal of Infectious Diseases (n=32). The first article related to this topic was published on January 17, 2020. The mean number of weekly publications in this cluster was 113.7 (SD 98.9), and the highest number of weekly publications was 266 in week 20 (Multimedia Appendix 1).

Topic 4: Epidemic Models for COVID-19 Spread

A total of 10.25% (2964/28,904) of the included publications were related to this topic (eg, [39-43]). The 5 most prominent authors in this cluster were Gerardo Chowell (n=22), Benjamin J Cowling (n=18), Kenji Mizumoto (n=14), Shi Zhao (n=13), and Rosalind M Eggo (n=13). The most common journals and preprint servers where the articles related to this topic were published included Chaos, Solitons & Fractals (n=73), medRxiv (n=66), the International Journal of Infectious Diseases (n=36), Zhonghua liuxingbingxue zazhi (n=30), and bioRxiv (n=26). The first paper related to this topic was published on the January 19, 2020. Although there was a sharp increase in the number of weekly publications between weeks 12 and 15, the trend was almost stable from week 15 to week 22. Then, a rapid decline in the number of weekly publications was noticed (Multimedia Appendix 1). The mean number of weekly publications in this cluster was 101.6 (SD 68.2).

Topic 5: Therapies and Vaccines for COVID-19

In all, 6.38% (1845/28,904) of the publications were about the development and repurposing of therapies and vaccines for COVID-19 (eg, [44-47]). The following authors published the highest number of articles related to this topic: Wei Zhang (n=13), Xiuna Yang (n=9), Haitao Yang (n=9), Zihe Rao (n=9), and Yao Zhao (n=8). The journals and preprint servers publishing the highest number of studies in this cluster were bioRxiv (n=174), the Journal of Biomolecular Structure and Dynamics (n=74), Trials (n=49), the Journal of Medical Virology (n=20), and Clinical Pharmacology & Therapeutics (n=14). In this cluster, the first article was published on January 6, 2020. The number of weekly publications increased dramatically from week 14 until a peak was reached in week 22 (n=144); thereafter, it decreased noticeably (Multimedia Appendix 1). The mean number of weekly publications in this cluster was 62.9 (SD 52.3).

Topic 6: Host Immune Response to 19-nCoV

This topic was discussed in about 6.36% (1837/28,904) of the publications (eg, [48-52]). Authors who had the highest number of publications related to this topic were Alessandro Sette (n=7), Stanley Perlman (n=6), Nima Rezaei (n=6), Irfan Rahman (n=5), and Akiko Iwasaki (n=6). The top 5 journals and preprint servers in terms of publishing articles related to this topic were bioRxiv (n=199), Medical Hypotheses (n=50), the Journal of Medical Virology (n=49), Frontiers in Immunology (n=22), and the British Journal of Haematology (n=19). The earliest article related to this topic was published on January 2, 2020. From that date until week 14, there was a slight increase in the number of weekly publications before it increased markedly, peaking in week 25 (n=155) (Multimedia Appendix 1). The mean number of weekly publications in this cluster was 62.8 (SD 54.6).

Topic 7: Diagnosis of COVID-19 Using Polymerase Chain Reaction

Using polymerase chain reaction (PCR) for diagnosing COVID-19 was a key topic discussed in 5.54% (1602/28,904) of the publications (eg, [53-57]). The most common authors writing about this topic were Alexander L Greninger (n=11), Kwok-Yung Yuen (n=10), Jasper Fuk-Woo Chan (n=9), Kelvin Kai-Wang To (n=9), and Cyril Chik-Yan Yip (n=9). The top 5 journals and preprint servers that published the highest number of studies related to this topic were bioRxiv (n=136), the Journal of Medical Virology (n=53), the Journal of Clinical Virology (n=40), medRxiv (n=31), and Clinical Infectious Diseases (n=30). The earliest study in this cluster was published at the beginning of week 3. The highest number of weekly publications was 119 in weeks 22 and 23 (Multimedia Appendix 1). The mean number of weekly publications related to this topic was 54.8 (SD 45.6).

Topic 8: Mental Health and Disorders During the COVID-19 Pandemic

This topic is about COVID-19–related mental health and disorders, which was explored by 3.17% (915/28,904) of the publications (eg, [58-62]). The top 5 authors in terms of number of publications related to this topic were Valerie A Canady (n=15), Mark D Griffiths (n=8), Stephen X Zhang (n=6), Zhilei Shang (n=5), and Modesto Leite Rolim Neto (n=5). The top 5 journals publishing studies related to this topic were Psychological Trauma: Theory, Research, Practice, and Policy (n=101); Psychiatry Research (n=48); the International Journal of Environmental Research and Public Health (n=37); the Journal of Affective Disorders (n=23); and Mental Health Weekly (n=23). In this cluster, the first article was published at the beginning of week 8. There was a considerable rise in the number of weekly publications from week 14 until a peak was reached in week 23 (n=94); this was followed by a steep decline (Multimedia Appendix 1). The mean number of weekly publications in this cluster was 31.1 (SD 30.9).

Topic 9: Diagnosis of COVID-19 Based on Chest Imaging

Diagnosis of COVID-19 based on chest imaging (eg, x-ray and computed tomography) was the main topic in 3.02% (874/28,904) of the included publications (eg, [63-67]). The 5 most prominent authors in this topic were Liming Xia (n=14), Michael Chung (n=12), Hongjun Li (n=11), Dinggang Shen (n=11), and Fuhua Yan (n=11). The top 5 journals in this cluster were European Radiology (n=38), Radiology (n=21), the American Journal of Roentgenology (n=17), Radiology: Cardiothoracic Imaging (n=16), and Clinical Nuclear Medicine (n=16). The first article about this topic was published on February 3, 2020. The mean number of weekly publications in this cluster was 29.9 (SD 21.4), and the highest number of weekly publications was 60 in week 18 (Multimedia Appendix 1).

Topic 10: Social Distancing Measures

A total of 3% (868/28,904) of the articles discussed the topic of social distancing measures used to fight against the COVID-19 pandemic (eg, [68-72]). Authors who had the highest number of publications related to this topic were Lei Zhang (n=7), Adam J Kucharski (n=6), Amy Gimma (n=5), Gerardo Chowell (n=5), and Petra Klepac (4). The top 5 journals and preprint servers in terms of publishing articles related to this topic were medRxiv (n=28); Chaos, Solitons & Fractals (n=6); Morbidity and Mortality Weekly Report (n=5); Science (n=5); and Disaster Medicine and Public Health Preparedness (n=5). The earliest article related to this topic was published in week 7. There was a dramatic rise in the number of weekly publications between week 12 and week 19; thereafter, the trend was unstable from week 20 to week 29 (Multimedia Appendix 1). The mean number of weekly publications in this cluster was 29.8 (SD 26.8).

Topic 11: Virus Genomics

Around 2.82% (816/28,904) of the publications were about genome sequences of 2019-nCoV (eg, [73-77]). The most common authors writing about this topic were Massimo Ciccozzi (n=11), Andrew Rambaut (n=11), Marta Giovanetti (n=10), Silvia Angeletti (n=10), and Domenico Benvenuto (n=9). The top 5 journals and preprint servers that published the highest number of studies in this cluster were bioRxiv (n=295); the Journal of Medical Virology (n=32); Microbiology Resource Announcements (n=14); Viruses (n=13); and Infection, Genetics and Evolution (n=12). The earliest study in this cluster was published at the beginning of week 4. The mean number of weekly publications in this cluster was 27.7 (SD 18.2). The highest number of weekly publications was 56 in week 24 (Multimedia Appendix 1).

Topic 12: Protein Structures of 2019-nCoV

About 2.44% (706/28,904) of the included publications focused on structures and functions of 2019-nCoV proteins (eg, [78-82]). The top 5 authors in terms of the number of publications related to this topic were Ralph S Baric (n=14), Jason S McLellan (n=12), Shibo Jiang (n=10), James Brett Case (n=10), and Daniel Wrapp (n=10). The journals and preprint servers publishing the highest number of studies in this cluster were bioRxiv (n=333), the Journal of Virology (n=15), the Journal of Biomolecular Structure & Dynamics (n=15), Science (n=13), and the Journal of Medical Virology (n=11). The earliest study related to this topic was published on January 3, 2020. The mean number of weekly publications in this cluster was 24.2 (SD 18.6). The highest number of weekly publications was 68 in week 25 (Multimedia Appendix 1).

Topic 13: Host Cell Entry

Host cell entry for 19-nCoV (via angiotensin-converting enzyme 2) was a key topic discussed in 2.02% (584/28,904) of the reviewed publications (eg, [83-87]). The 5 most prominent authors in this cluster were Serpil Erzurum (n=4), Giuseppe Lippi (n=4), Daniel Batlle (n=4), Hong Gao (n=4), and Claudio Cavallini (n=3). The most common journals and preprint servers in this cluster were bioRxiv (n=117), the Journal of Medical Virology (n=11), Medical Hypotheses (n=10), European Respiratory Journal (n=8), and medRxiv (n=7). The first article related to this topic was published in the mid of week 4. The number of weekly publications was almost stable between weeks 4 and 13. Thereafter, a sharp increase was noticed between weeks 14 and 16, but it was not stable from then until week 29 (Multimedia Appendix 1). The number of publications in week 20 was the highest (n=50). The mean number of weekly publications in this cluster was 19.9 (SD 16.6).

Topic 14: Patients With Cancer During the COVID-19 Pandemic

A total of 1.53% (441/28,904) of the included publications were about patients with cancer during the COVID-19 pandemic (eg, [88-91]). The following authors published the highest number of articles related to this topic: Solange Peters (n=6), Umberto Ricardi (n=5), Conghua Xie (n=5), Giuseppe Curigliano (n=5), and Alessio Cortellini (n=4). The top 5 journals in terms of publishing articles related to this topic were Head & Neck (n=16), ecancermedicalscience (n=15), Radiotherapy and Oncology (n=10), Advances in Radiation Oncology (n=8), and Cancer Discovery (n=8). From week 1 to 7, only 1 study related to this topic was published. The highest number of weekly

publications was in week 22 (n=49), but since then, the number of weekly publications declined steeply (Multimedia Appendix 1). The mean number of weekly publications related to this topic was 15.2 (SD 14.4).

Topic 15: Detection of 2019-nCoV Antibodies

Detection of antibodies against 2019-nCoV using serological assays was a topic discussed in 1.42% (411/28,904) of all publications (eg, [92-96]). The top 5 authors writing about this topic were Florian Krammer (n=6), Jing Wang (n=6), Yong Zhang (n=5), Juan Chen (n=5), and Viviana Simon (n=5). The top 5 journals and preprint servers that published the highest number of studies in this cluster were bioRxiv (n=30), medRxiv (n=20), the Journal of Medical Virology (n=18), the Journal of Clinical Virology (n=14), and the Journal of Clinical Microbiology (n=7). Only 1 study in this cluster was published in the first 6 weeks. There was a dramatic increase in the number of weekly publications between weeks 16 and 21. Although the number of weekly publications slightly decreased from week 22 until week 26, it increased rapidly until reaching the peak in weeks 28 and 29 (n=36) (Multimedia Appendix 1). The mean number of weekly publications in this cluster was 14.1 (SD 12.8).

Topic 16: Personal Protective Equipment

Around 1.21% (350/28,904) of the publications focused on personal protective equipment in the COVID-19 era (eg, [97-100]). Authors who published the most in this cluster were Holly Seale (n=3), Keith K Wannomae (n=3), Lei Liao (n=3), Wang Xiao (n=3), and Steven Chu (n=3). The highest numbers of studies were published in the following journals and preprint servers: the American Journal of Infection Control (n=8), the Journal of Hospital Infection (n=7), Anaesthesia (n=6), the Journal of the European Academy of Dermatology and Venereology (n=6), ACS Nano (n=5), and medRxiv (n=5). No articles in this cluster were published before week 9. The mean number of weekly publications in this cluster was 11.9 (SD 11.6). The highest number of weekly publications was 34 in week 23 (Multimedia Appendix 1).

Topic 17: Diabetes Mellitus and COVID-19

Health care management, clinical characteristics, and risk factors for mortality of COVID-19 patients with diabetes was discussed in 1.16% (336/28,904) of the included articles (eg, [101-104]. The 5 most prominent authors in this cluster were Hui Wang (n=5), Sam Foster (n=4), Anoop Misra (n=4), Béatrice Bouhanick (n=3), and Kamlesh Khunti (n=3). The most common journals in this cluster were Diabetes Research and Clinical Practice (n=20), Diabetology & Metabolic Syndrome (n=17), the British Journal of Nursing (n=10), the Journal of the American Medical Directors Association (n=7), and Diabetes Technology & Therapeutics (n=6). Only 4 articles related to this topic were published between weeks 1 and 12. However, there was a substantial increase in the number of weekly publications from week 17 until the peak was reached in week 20 (n=35); this was followed by a slight decrease (Multimedia Appendix 1). The mean number of weekly publications in this cluster was 11.4 (SD 11.6).

Topic 18: Pregnancy and Childbirth During the COVID-19 Pandemic

About 1.08% (312/28,904) of the publications focused on numerous aspects of pregnancy and childbirth during the COVID-19 pandemic (eg, [105-109]). The most common authors writing about this topic were Ling Feng (n=7), Jiafu Li (n=6), Olivier Picone (n=5), Dunjin Chen (n=5), and Guoqiang Sun (n=5). The top 5 journals in terms of publishing articles in this cluster were the International Journal of Gynaecology and Obstetrics (n=18), The Journal of Maternal-Fetal & Neonatal Medicine (n=16), the American Journal of Obstetrics and Gynecology (n=20), Obstetrics and Gynecology (n=10), and the American Journal of Perinatology (n=9). The earliest article in this cluster was published on February 10, 2020. The mean number of weekly publications related to this topic was 10.8 (SD 8.9), and the highest number of weekly publications was 27 in week 21 (Multimedia Appendix 1).

Topic 19: Organ Transplantation During the COVID-19 Pandemic

Organ transplantation in the era of COVID-19 was a key topic in 0.76% (219/28,904) of the included articles (eg, [110-113]) The top 5 authors in terms of number of publications related to this topic were Paolo Cravedi (n=4), Zhishui Chen (n=4), Luciano De Carlis (n=4), Lai Wei (n=4), and Ashley Fan (n=3). The top 5 journals in terms of publishing articles related to this topic were the American Journal of Transplantation (n=69), Transplant Infectious Disease (n=35), Transplant International (n=11), Transplantation Proceedings (n=10), and Liver Transplantation (n=5). Only one study in this cluster was published before week 12. The mean number of weekly publications related to this topic was 7.5 (SD 8.3), and the highest number of weekly publications was 26 in week 24 (Multimedia Appendix 1).

Discussion

Principal Findings

We found that 5.92% (1714/28,904) of the included published articles were hosted on preprint servers (bioRxiv or medRxiv). Although these servers are not the only preprint servers available in the academic publishing landscape (many journals publish articles online before they go into print, and we have also observed a rise of purely online journals), they are indicative of the pace with which new knowledge is made available by the international research community. Since such preprint servers do not undergo formal peer reviewing and are, thus, not regarded publications in the traditional academic sense, many researchers are using this device to make findings available and to solicit feedback from the international community before undergoing formal peer-reviewing by journals—a process that takes at least 2 months to get the submitted paper published.

Among the peer-reviewed journals, the Journal of Medical Virology has published the highest number of COVID-19–related articles (n=468). Aristovnik et al and Hossain also listed the Journal of Medical Virology in the top-5 journals publishing COVID-19–related articles [6,14]. The Journal of Medical Virology clearly stands out, as it has

published more than twice the number of papers compared to the second-ranked journal—the International Journal of Environmental Research and Public Health (n=223). Aristovnik et al [6] listed the International Journal of Environmental Research and Public Health among the 10 top-ranked journals based on COVID-19–related research articles [6]. The source normalized impact per paper (SNIP), in the year 2019, was 0.780 for the Journal of Medical Virology [114] and 1.248 for the International Journal of Environmental Research and Public Health [115], and the average time from the submission to the first decision was about 6 weeks [116] and 3 weeks [117], respectively. We believe the speed of the reviewing process of these journals may have motivated the authors to submit their work to these journals.

Considering the study methods, we found that the highest number of studies (n=1515) were surveys, followed by reviews (systematic review, scoping review, or meta-analyses), as shown in Table 3. As the number of research studies on COVID-19 is rapidly increasing, review articles are of utmost importance to summarize the ongoing effort and progress to combat against COVID-19. We found case-control studies to be the lowest represented study design (n=62 only). We speculate that the lack of available data was the main reason for the scarcity of this type of research study. Interestingly, 362 randomized control trials in 7 months indicate the enormous effort made by the scientific community to combat this pandemic.

Furthermore, we grouped the 19 topics addressed in the included studies into six thematic areas (summarized in Table 6). The dominant thematic clusters were “Clinical aspects” (29.17%) and “Epidemiology” (28.91%). The “Clinical aspects” theme covers multiple aspects of the clinical practices for patient care and risk factors related to COVID-19. It consists of two topics (ie, “clinical care practices for patients during the COVID-19 pandemic” and “clinical characteristics and risk factors of COVID-19). Interestingly, the “Epidemiology” theme also comprises only two topics (ie, “Public health response” and “Epidemic models for COVID-19 spread”), further underscoring the dominance of these topics.

Table 6. Topics grouped by thematic cluster, including the percentage of articles by topic and cluster.

Thematic cluster, topic number, and title		Articles (N=28,904), n	Topic dominance (%)	Cluster dominance (%)
Clinical aspects				29.17
(2)	Clinical care practices for patients during the COVID-19 pandemic	5118	17.71	
(3)	Clinical characteristics and risk factors of COVID-19	3313	11.46	
Epidemiology				28.91
(1)	Public health response	5393	18.66	
(4)	Epidemic models for COVID-19 spread	2964	10.25	
Therapeutics				21.03
(5)	Therapies and vaccines for COVID-19	1845	6.38	
(6)	Host immune response	1837	6.36	
(11)	Virus genomics	816	2.82	
(12)	Protein structures of 2019-nCoV	706	2.44	
(13)	Host cell entry	584	2.02	
Diagnostics				9.98
(7)	Diagnosis of COVID-19 using PCR	1602	5.54	
(9)	Diagnosis of COVID-19 based on chest imaging	874	3.02	
(15)	Detection of 2019-nCoV antibodies	411	1.42	
Related conditions				7.70
(8)	Mental health and disorders during the COVID-19 pandemic	915	3.17	
(14)	Patients with cancer during the COVID-19 pandemic	441	1.53	
(17)	Diabetes mellitus and COVID-19	336	1.16	
(18)	Pregnancy and childbirth during the COVID-19 pandemic	312	1.08	
(19)	Organ transplantation during the COVID-19 pandemic	219	0.76	
Prevention				4.21
(10)	Social distancing measures	868	3	
(16)	Personal protective equipment	350	1.21	

The third most prominent theme "Therapeutics" (21.03%) comprises five topics, making it the most diverse theme; the topics in this theme range from "host cell entry" to drug discovery–related terms such as "Protein structures of 2019-nCoV" and "Virus genomics," as well as "Therapies and vaccines for COVID-19." This theme highlights the initiatives of the scientific community to discover drugs and vaccines and understand the underlying virus-host mechanism to pave the way for effective therapeutic solutions for COVID-19. Considering the severity of COVID-19, we believe there still may be a lack of publications in this theme, despite comprising slightly more than 20% of all articles. We believe that, as clinical practices and public health responses mature, this theme will receive more research articles in the near future.

Almost 10% of articles form the "Diagnostics" theme. This theme focuses on the diagnosis of COVID-19 based on PCR, radiological images, or antibodies. PCR is among the most accurate technologies to diagnose COVID-19 [114], which explains the numerous relevant publications. Due to the advancement of deep learning techniques, radiological image-based diagnosis is becoming more effective and has the potential to save time in clinical environments [115]. As a result, we observed a large number of publications on radiological image-based analyses, which is captured under the topic "Diagnosis of COVID-19 based on chest imaging." Antibodies, developed in hosts combatting the novel coronavirus, can be considered a detection mechanism that may play an important role complementary to PCR testing [116]. It can be very effective for the diagnosis of patients with asymptomatic COVID-19 or negative RT–PCR results [117]. We also noticed many publications on antibody responses against COVID-19, which are covered by the topic "Detection of 2019-nCoV antibodies".

The interplay between COVID-19 and related medical conditions is captured by the "Related conditions" theme. This theme not only comprises articles discussing related conditions caused by COVID-19 but also other conditions that may elevate the COVID-19 risk for the patients with those conditions. About 8% of all articles fall into this theme, covering topics such as mental disorder, diabetes, cancer, pregnancy, childbirth complications, and organ transplantation.

Only slightly more than 4% of all articles fall into the "Prevention" theme. This may be surprising, since prevention is of utmost importance while vaccines and treatments are still under development. However, we believe that this theme is not covered by more studies due to the recent wide acceptance and effectiveness of social distancing and personal protective equipment. Consequently, we expect the percentage of articles grouped in this theme to further reduce in the future. Further insights into the research landscape and the shift in themes over time is summarized in Multimedia Appendix 2.

We noticed that biomedical informatics had a crucial role in several topics. For instance, clinical decision support systems were used in many studies to diagnose COVID-19 based on chest imaging. Telemedicine was also used in multiple studies to provide the required health care support for the patients during the COVID-19 pandemic. Further, mobile applications, including contact tracing apps, were one of the main social distancing measures described in these studies. AI-based models were used in multiple studies to predict protein structures of 2019-nCoV to understand the underlying mechanism of drug-target interaction. Many studies proposed novel AI-based models to discover COVID-19 drugs and vaccines and repurpose existing drugs approved by the Food and Drug Administration as a part of the treatment plan for COVID-19.

Strengths and Limitations

Strengths

To the best of our knowledge, our study covers the largest collection of COVID-19–related articles (N=28,904, after considering the inclusion and exclusion criteria) published in the period of 7 months (January to mid-July 2020). The main strength of this study is that it demonstrates the feasibility of mostly automated, AI-based data mining at scale. We believe this is of utmost importance because articles on COVID-19 are published faster than nonautomatic surveys can organize, analyze, and present them to the scientific community.

Limitations

All the publications were collected from only one specific database (CORD-19), so we may have missed some studies or preprints that were not considered in this database. Given the substantial number of publications included in our analysis, we are confident that a large part of the COVID-19 literature was covered. Further, we did not conduct a detailed manual analysis of studies published in journals such as Journal of Medical Internet Research and the International Journal of Medical Informatics to evaluate the use of eHealth technologies for COVID-19, which we believe is beyond the scope of our work and requires other study methodologies such as systematic or scoping review. However, readers are referred to several studies and reviews, which have been conducted to explore eHealth technologies used in the fight against the COVID-19 pandemic [118-121].

We only considered the articles published in the English language, which may introduce some bias in our analysis. Additionally, articles published after mid-July 2020 were not considered in this study. Moreover, due to the inherent limitation of the bibliographic analysis, which enabled high-level profiling of the text from the corpus of literature, we cannot provide any evidence-based solution for the diagnosis and treatment of COVID-19. Further analysis of the articles that fall under each topic should be considered more carefully.

For this study, we only analyzed article titles, abstracts, and author data. As a result, we could not identify the country of publication for 11,634 articles (around 40%). Although we removed duplicates, academic publications undergo subtle morphology changes. These come in the form of revisions, preprints, follow-up studies, among others. Given only the abstract and title were screened, we cannot rule out that some publications may have substantial overlaps. Finally, there exist ambiguities with respect to author and journal names because authors with the same name can only be resolved uniquely by affiliation or, in some cases, other identifiers such as ORCID. If such elucidating identifiers are missing, automatic

disambiguation is not possible. Likewise, journals may be referred to by a plethora of acronyms. Therefore, our study may have merged multiple authors into one person and split journals into multiple entities.

Practical and Research Implications

Practical Implications

This study demonstrates the feasibility of AI-based, largely automatic data mining of large corpora of academic publications. Given the pace and dynamics at which COVID-19 research is being conducted at this time, attempts to manually survey the literature will almost certainly fall behind the state-of-the-art, *unless* specific and confined subtopics are under scrutiny. Automatic data mining, on the other hand, is hindered by the inherent noise in the data (eg, lack of ORCIDs and inability to track genesis and evolution of articles properly). This means that results accurate to every single paper and author are impossible to extract, unless publishers (eg, using blockchain technology) develop a way to trace genesis and evolution of individual publications and unique authors. It is therefore important to stress that such automatic textual analysis is performed *at scale*. If conducted at scale, we would argue that exact numbers may not matter as much as they once may have, considering that being off by as many as 30 individual publications while screening of almost 29,000 publications means that being off by as many as 30 individual publications is would still constitute a negligible fraction of the overall corpus. We, therefore, believe the numbers we have presented in this report to be *aggregates* that are representative of a vibrant and rapidly evolving research landscape and that they highlight trends and shifting interests in topics. Whereas automated data mining excels at providing up-to-date, broad overviews of the field as a whole, manual surveys excel at providing detailed overviews of specific topics. We, therefore, see this study complementing previous manual reviews.

Research Implications

This study highlighted the effectiveness of AI methods in the analysis of a large corpus of literature, which researchers can use to perform machine learning–based bibliometric analysis of eHealth-related literature to explore the use of eHealth technologies for COVID-19.

Among the research themes summarized in Table 6, we see the direst need for more research in the “Therapeutics” theme, as clinical aspects and epidemiological aspects are better understood and best practices continue to be more commonly implemented. Consequently, we see this shift in the proportion of articles at the expense of the “Clinical Aspects” and “Epidemiology” themes, as well as the “Prevention” theme. The reason is that the impact of related topics “Social distancing” and “Personal preventive equipment” should be well understood and implemented by now.

In the context of performing bibliometric reviews based on automatically extracted topics, the most important research challenge is to develop methods that are more robust against noise and can process not only abstracts, titles, and author lists, but the entire full text of the publications. Although computationally extremely demanding, this would allow assessing any overlaps between publications, which is a stronger measure than the binary decision of duplicity.

Conclusions

This study provides a comprehensive overview of the COVID-19 literature. Specifically, we identified the main COVID-19–related topics addressed in the existing literature; weekly trends of publications; and top countries, authors, and publishers. This study will help the research community to understand the evolution of the COVID-19–related literature; prioritize research needs; and recognize the leading researchers, institutes, countries, and publishers for each topic. AI-based bibliometric analysis has the potential to rapidly explore large corpora of academic publications. Publishers should avoid noise in the data by developing a way to trace the evolution of individual publications and unique authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Number of publications for each topic.
[DOCX File , 102 KB - jmir_v23i3e23703_app1.docx]

Multimedia Appendix 2
Number of publications for each theme.
[DOCX File , 43 KB - jmir_v23i3e23703_app2.docx]

References

1. Novel Coronavirus (2019-nCoV) Situation Report-1. World Health Organization. Geneva, Switzerland; 2020 Jan 21. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4 [accessed 2020-07-19]

2. Chen Y, Li L. SARS-CoV-2: virus dynamics and host response. Lancet Infect Dis 2020 May;20(5):515-516 [FREE Full text] [doi: 10.1016/S1473-3099(20)30235-8] [Medline: 32213336]
3. Fontanarosa PB, Bauchner H. COVID-19--looking beyond tomorrow for health care and society. JAMA 2020 May 19;323(19):1907-1908. [doi: 10.1001/jama.2020.6582] [Medline: 32301955]
4. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. Geneva, Switzerland; 2020. URL: https://covid19.who.int/ [accessed 2020-08-20]
5. Yang P, Wang X. COVID-19: a new challenge for human beings. Cell Mol Immunol 2020 May 31;17(5):555-557 [FREE Full text] [doi: 10.1038/s41423-020-0407-x] [Medline: 32235915]
6. Aristovnik A, Rav?elj D, Umek L. A bibliometric analysis of COVID-19 across science and social science research landscape. Preprints. Preprint published online on June 24, 2020. [doi: 10.20944/preprints202006.0299.v1]
7. Haghani M, Bliemer M, Goerlandt F, Li J. The scientific literature on Coronaviruses, COVID-19 and its associated safety-related research dimensions: A scientometric analysis and scoping review. Saf Sci 2020 Sep;129:104806 [FREE Full text] [doi: 10.1016/j.ssci.2020.104806] [Medline: 32382213]
8. Doanvo A, Qian X, Ramjee D, Piontkivska H, Desai A, Majumder M. Machine learning maps research needs in COVID-19 literature. bioRxiv. Preprint posted online on June 5, 2020. [doi: 10.1101/2020.06.11.145425]
9. Global Coronavirus COVID-19 Clinical Trial Tracker. Cytel. URL: https://www.covid-trials.org/ [accessed 2020-07-13]
10. Dong M, Cao X, Liang M, Li L, Liang H, Liu G. Understand research hotspots surrounding COVID-19 and other coronavirus infections using topic modeling. medRxiv. Preprint posted online on April 10, 2020. [doi: 10.1101/2020.03.26.20044164]
11. Le BP, Gharavi A, Robb DA, Vidal AF, Padilla S, Chantler MJ. Visualising COVID-19 research. arXiv. Preprint posted online on May 12, 2020 [FREE Full text]
12. Mao X, Guo L, Fu P, Xiang C. The status and trends of coronavirus research: a global bibliometric and visualized analysis. Medicine (Baltimore) 2020 May 29;99(22):e20137 [FREE Full text] [doi: 10.1097/MD.0000000000020137] [Medline: 32481379]
13. Chahrour M, Assi S, Bejjani M, Nasrallah AA, Salhab H, Fares M, et al. A bibliometric analysis of COVID-19 research activity: a call for increased output. Cureus 2020 Mar 21;12(3):e7357 [FREE Full text] [doi: 10.7759/cureus.7357] [Medline: 32328369]
14. Hossain MM. Current status of global research on novel coronavirus disease (COVID-19): a bibliometric analysis and knowledge mapping. F1000Res 2020 May 18;9:374. [doi: 10.12688/f1000research.23690.1]
15. Lou J, Tian SJ, Niu SM, Kang XQ, Lian HX, Zhang LX, et al. Coronavirus disease 2019: a bibliometric analysis and review. Eur Rev Med Pharmacol Sci 2020 Mar;24(6):3411-3421 [FREE Full text] [doi: 10.26355/eurrev_202003_20712] [Medline: 32271460]
16. Nasab R, Rahim F. Bibliometric analysis of global scientific research on SARSCoV-2 (COVID-19). medRxiv. Preprint posted online on March 23, 2020. [doi: 10.1101/2020.03.19.20038752]
17. Ahamed S, Samad M. Information mining for COVID-19 research from a large volume of scientific literature. arXiv. Preprint posted online April 5, 2020 [FREE Full text]
18. Bonilla-Aldana DK, Quintero-Rada K, Montoya-Posada JP, Ramírez-Ocampo S, Paniz-Mondolfi A, Rabaan AA, et al. SARS-CoV, MERS-CoV and now the 2019-novel CoV: Have we investigated enough about coronaviruses? - A bibliometric analysis. Travel Med Infect Dis 2020 Jan;33:101566 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101566] [Medline: 32007621]
19. Aguado-Cortés C, Castaño V. Translational knowledge map of COVID-19. arXiv. Preprint posted online on March 22, 2020.
20. CORD-19: COVID-19 Open Research Dataset. Allen Institute for AI. URL: https://allenai.org/data/cord-19 [accessed 2020-07-07]
21. COVID-19-Literature-Mining Repository. GitHub. URL: https://github.com/zshah-HBKU/COVID-19-Literature-Mining- [accessed 2021-02-24]
22. Aggarwal C, Zhai C. A Survey of Text Classification Algorithms. In: Aggarwal C, Zhai C, editors. Mining Text Data. Boston, MA: Springer; 2012.
23. Marutho D, Handaka S, Wijaya E. The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News. : IEEE; 2018 Presented at: 2018 International Seminar on Application for Technology of Information and Communication; September 21-22, 2018; Semarang, Indonesia p. 533-538 URL: https://ieeexplore.ieee.org/document/8549751 [doi: 10.1109/isemantic.2018.8549751]
24. AlQutob R, Moonesar IA, Tarawneh MR, Al Nsour M, Khader Y. Public health strategies for the gradual lifting of the public sector lockdown in Jordan and the United Arab Emirates during the COVID-19 crisis. JMIR Public Health Surveill 2020 Jul 21;6(3):e20478 [FREE Full text] [doi: 10.2196/20478] [Medline: 32589151]
25. Bell D, Schultz HK, Kiragga A, Kambugu A, Kissa J, Mbonye A. COVID-19 in Uganda: predicting the impact of the disease and public health response on disease burden. medRxiv. Preprint posted online on June 10, 2020. [doi: 10.1101/2020.05.14.20102202]
26. Bouey J. Strengthening China's public health response system: from SARS to COVID-19. Am J Public Health 2020 Jul;110(7):939-940. [doi: 10.2105/AJPH.2020.305654] [Medline: 32213081]

27. Deng S, Peng H. Characteristics of and public health responses to the coronavirus disease 2019 outbreak in China. J Clin Med 2020 Feb 20;9(2):575 [FREE Full text] [doi: 10.3390/jcm9020575] [Medline: 32093211]
28. Maher PJ, MacCarron P, Quayle M. Mapping public health responses with attitude networks: the emergence of opinion-based groups in the UK's early COVID-19 response phase. Br J Soc Psychol 2020 Jul;59(3):641-652 [FREE Full text] [doi: 10.1111/bjso.12396] [Medline: 32621294]
29. Bajunaid K, Alqurashi A, Alatar A, Alkutbi M, Alzahrani A, Sabbagh A, et al. Neurosurgical procedures and safety during the COVID-19 pandemic: a case-control multicenter study. World Neurosurg. Epub ahead of print posted online on July 20, 2020 2020 [FREE Full text] [doi: 10.1016/j.wneu.2020.07.093]
30. Qazi SH, Saleem A, Pirzada AN, Hamid L, Dogar SA, Das JK. Challenges to delivering pediatric surgery services in the midst of COVID 19 crisis: experience from a tertiary care hospital of Pakistan. Pediatr Surg Int. Epub ahead of print posted online on July 19, 2020 2020 [FREE Full text] [doi: 10.1007/s00383-020-04721-0]
31. Rinkel LA, Prick JCM, Slot RER, Sombroek NMA, Burggraaff J, Groot AE, et al. Impact of the COVID-19 outbreak on acute stroke care. J Neurol. Epub ahead of print posted online on July 19, 2020 2020 [FREE Full text] [doi: 10.1007/s00415-020-10069-1]
32. Hajdu SD, Pittet V, Puccinelli F, Ben Hassen W, Ben Maacha M, Blanc R, et al. Acute stroke management during the COVID-19 pandemic. Stroke. Epub ahead of print posted online on July 1, 2020 2020. [doi: 10.1161/strokeaha.120.030794]
33. Limoncelli J, Marino T, Smetana R, Sanchez-Barranco P, Brous M, Cantwell K, et al. General anesthesia recommendations for electroconvulsive therapy during the coronavirus disease 2019 pandemic. J ECT 2020 Sep;36(3):152-155 [FREE Full text] [doi: 10.1097/YCT.0000000000000705] [Medline: 32483035]
34. He Y, Lian S, Dong Y. Clinical characteristics, diagnosis, and treatment of COVID-19: A case report. World J Clin Cases 2020 Jun 06;8(11):2325-2331 [FREE Full text] [doi: 10.12998/wjcc.v8.i11.2325] [Medline: 32548163]
35. Gémes K, Talbäck M, Modig K, Ahlbom A, Berglund A, Feychting M, et al. Burden and prevalence of prognostic factors for severe COVID-19 in Sweden. Eur J Epidemiol 2020 May;35(5):401-409 [FREE Full text] [doi: 10.1007/s10654-020-00646-z] [Medline: 32424571]
36. Sharma R, Agarwal M, Gupta M, Somendra S, Saxena S. Clinical characteristics and differential clinical diagnosis of novel coronavirus disease 2019 (COVID-19). In: Saxena S, editor. Coronavirus Disease 2019 (COVID-19). Medical Virology: From Pathogenesis to Disease Control. Singapore: Springer; 2020:55-70.
37. Ntaios G, Michel P, Georgiopoulos G, Guo Y, Li W, Xiong J, et al. Characteristics and outcomes in patients with COVID-19 and acute ischemic stroke: The Global COVID-19 Stroke Registry. Stroke. Epub ahead of print posted online on July 9, 2020 2020 [FREE Full text] [doi: 10.1161/STROKEAHA.120.031208]
38. Krishnan S, Patel K, Desai R, Sule A, Paik P, Miller A, et al. Clinical comorbidities, characteristics, and outcomes of mechanically ventilated patients in the State of Michigan with SARS-CoV-2 pneumonia. J Clin Anesth. Epub ahead of print posted online on July 20, 2020 2020 [FREE Full text] [doi: 10.1016/j.jclinane.2020.110005]
39. Endo A, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Res 2020 Jul 3;5:67. [doi: 10.12688/wellcomeopenres.15842.2]
40. Huang R, Liu M, Ding Y. Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis. J Infect Dev Ctries 2020 Mar;14(3):246-253 [FREE Full text] [doi: 10.3855/jidc.12585] [Medline: 32235084]
41. Roy S, Bhunia GS, Shit PK. Spatial prediction of COVID-19 epidemic using ARIMA techniques in India. Model Earth Syst Environ 2020 Jul 16:1-7. [doi: 10.1007/s40808-020-00890-y] [Medline: 32838022]
42. Zeb A, Alzahrani E, Erturk VS, Zaman G. Mathematical model for coronavirus disease 2019 (COVID-19) containing isolation class. Biomed Res Int 2020 Jun;2020:3452402 [FREE Full text] [doi: 10.1155/2020/3452402] [Medline: 32685469]
43. Karako K, Song P, Chen Y, Tang W. Analysis of COVID-19 infection spread in Japan based on stochastic transition model. Biosci Trends 2020 May 21;14(2):134-138 [FREE Full text] [doi: 10.5582/bst.2020.01482] [Medline: 32188819]
44. Checcucci E, Piramide F, Pecoraro A, Amparore D, Campi R, Fiori C, et al. The vaccine journey for COVID-19: a comprehensive systematic review of current clinical trials in humans. Panminerva Med. Epub ahead of print posted online on May 26, 2020 2020 [FREE Full text] [doi: 10.23736/S0031-0808.20.03958-0] [Medline: 32456404]
45. Koirala A, Joo Y, Khatami A, Chiu C, Britton P. Vaccines for COVID-19: The current state of play. Paediatr Respir Rev. Epub ahead of print posted online on June 18, 2020 2020 [FREE Full text] [doi: 10.1016/j.prrv.2020.06.010]
46. Mitjà O, Corbacho-Monné M, Ubals M, Tebe C, Peñafiel J, Tobias A, BCN PEP-CoV-2 RESEARCH GROUP. Hydroxychloroquine for early treatment of adults with mild Covid-19: a randomized-controlled trial. Clin Infect Dis 2020 Jul 16 [FREE Full text] [doi: 10.1093/cid/ciaa1009] [Medline: 32674126]
47. Willis V, Arriaga Y, Weeraratne D, Reyes F, Jackson G. A narrative review of emerging therapeutics for COVID-19. Mayo Clin Proc Innov Qual Outcomes. Epub ahead of print posted online on July 20, 2020 2020 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2020.07.004]
48. Chen J, Kelley WJ, Goldstein DR. Role of aging and the immune response to respiratory viral infections: potential implications for COVID-19. J Immunol 2020 Jul 15;205(2):313-320. [doi: 10.4049/jimmunol.2000380] [Medline: 32493812]
49. Gharote MA. Role of poly (ADP) ribose polymerase-1 inhibition by nicotinamide as a possible additive treatment to modulate host immune response and prevention of cytokine storm in COVID-19. IJMS 2020 Apr 30;72:25-28. [doi: 10.25259/ijms_29_2020]

50. Wilk AJ, Rustagi A, Zhao NQ, Roque J, Martínez-Colón GJ, McKechnie JL, et al. A single-cell atlas of the peripheral immune response in patients with severe COVID-19. Nat Med 2020 Jul;26(7):1070-1076 [FREE Full text] [doi: 10.1038/s41591-020-0944-y] [Medline: 32514174]
51. Zinovkin RA, Grebenchikov OA. Transcription factor Nrf2 as a potential therapeutic target for prevention of cytokine storm in COVID-19 patients. Biochemistry Moscow 2020 Jul 13;85(7):833-837. [doi: 10.1134/s0006297920070111]
52. Kumar S, Nyodu R, Maurya VK, Saxena SK. Host Immune Response and Immunobiology of Human SARS-CoV-2 Infection. In: Saxena S, editor. Coronavirus Disease 2019 (COVID-19). Medical Virology: From Pathogenesis to Disease Control. Singapore: Springer; Apr 2020:43-53.
53. Di Paolo M, Iacovelli A, Olmati F, Menichini I, Oliva A, Carnevalini M, et al. False-negative RT-PCR in SARS-CoV-2 disease: experience from an Italian COVID-19 unit. ERJ Open Res 2020 Apr;6(2) [FREE Full text] [doi: 10.1183/23120541.00324-2020] [Medline: 32685435]
54. Martínez-Murcia A, Bru G, Navarro A, Ros-Tárraga P, García-Sirera A, Pérez L. Comparative in silico design and validation of GPS™ CoVID-19 dtec-RT-qPCR test. J Appl Microbiol. Epub ahead of print posted online on July 29, 2020 2020 [FREE Full text] [doi: 10.1111/jam.14781]
55. Nagura-Ikeda M, Imai K, Tabata S, Miyoshi K, Murahara N, Mizuno T. Clinical evaluation of self-collected saliva by RT-qPCR, direct RT-qPCR, RT-LAMP, and a rapid antigen test to diagnose COVID-19. medRxiv. Preprint posted online on June 08, 2020. [doi: 10.1101/2020.06.06.20124123]
56. Nalla AK, Casto AM, Huang MW, Perchetti GA, Sampoleo R, Shrestha L, et al. Comparative performance of SARS-CoV-2 detection assays using seven different primer-probe sets and one assay kit. J Clin Microbiol 2020 May;58(6):e00557-e00520 [FREE Full text] [doi: 10.1128/jcm.00557-20]
57. Visseaux B, Le Hingrat Q, Collin G, Bouzid D, Lebourgeois S, Le Pluart D, et al. Evaluation of the QIAstat-Dx Respiratory SARS-CoV-2 Panel, the first rapid multiplex PCR commercial assay for SARS-CoV-2 detection. J Clin Microbiol 2020 Jul 23;58(8):e00630-e00620. [doi: 10.1128/jcm.00630-20]
58. Badahdah A, Khamis F, Al Mahyijari N, Al Balushi M, Al Hatmi H, Al Salmi I, et al. The mental health of health care workers in Oman during the COVID-19 pandemic. Int J Soc Psychiatry 2020 Jul 08:20764020939596 [FREE Full text] [doi: 10.1177/0020764020939596] [Medline: 32635837]
59. Cameron E, Joyce K, Delaquis C, Reynolds K, Protudjer J, Roos L. PsyArXiv Preprints Preprint posted online May 01, 2020. [FREE Full text] [doi: 10.31234/osf.io/a53zb]
60. Shaukat N, Ali DM, Razzak J. Physical and mental health impacts of COVID-19 on healthcare workers: a scoping review. Int J Emerg Med 2020 Jul 20;13(1):40 [FREE Full text] [doi: 10.1186/s12245-020-00299-5] [Medline: 32689925]
61. Xiaoming X, Ming A, Su H, Wo W, Jianmei C, Qi Z, et al. The psychological status of 8817 hospital workers during COVID-19 epidemic: a cross-sectional study in Chongqing. J Affect Disord. Epub ahead of print posted online on July 19, 2020 2020 [FREE Full text] [doi: 10.1016/j.jad.2020.07.092] [Medline: 32871686]
62. Zhao H, He X, Fan G, Li L, Huang Q, Qiu Q, et al. COVID-19 infection outbreak increases anxiety level of general public in China: involved mechanisms and influencing factors. J Affect Disord. Epub ahead of print posted online on July 20, 2020 2020 [FREE Full text] [doi: 10.1016/j.jad.2020.07.085]
63. Dangis A, Gieraerts C, De Bruecker Y, Janssen L, Valgaeren H, Obbels D, et al. Accuracy and reproducibility of low-dose submillisievert chest CT for the diagnosis of COVID-19. Radiology: Cardiothoracic Imaging 2020 Apr 01;2(2):e200196. [doi: 10.1148/ryct.2020200196]
64. Fang M, He B, Li L, Dong D, Yang X, Li C, et al. CT radiomics can help screen the coronavirus disease 2019 (COVID-19): a preliminary study. Sci. China Inf. Sci 2020 Apr 15;63(7):172103. [doi: 10.1007/s11432-020-2849-3]
65. Wen Z, Chi Y, Zhang L, Liu H, Du K, Li Z, et al. Coronavirus disease 2019: initial detection on chest CT in a retrospective multicenter study of 103 Chinese patients. Radiology: Cardiothoracic Imaging 2020 Apr 01;2(2):e200092. [doi: 10.1148/ryct.2020200092]
66. Yang R, Li X, Liu H, Zhen Y, Zhang X, Xiong Q, et al. Chest CT severity score: an imaging tool for assessing severe COVID-19. Radiology: Cardiothoracic Imaging 2020 Apr 01;2(2):e200047. [doi: 10.1148/ryct.2020200047]
67. Yu M, Xu D, Lan L, Tu M, Liao R, Cai S, et al. Thin-section chest CT imaging of COVID-19 pneumonia: a comparison between patients with mild and severe disease. Radiology: Cardiothoracic Imaging 2020 Apr 01;2(2):e200126. [doi: 10.1148/ryct.2020200126]
68. Courtemanche C, Garuccio J, Le A, Pinkston J, Yelowitz A. Strong social distancing measures in the United States reduced the COVID-19 growth rate. Health Aff (Millwood) 2020 Jul;39(7):1237-1246. [doi: 10.1377/hlthaff.2020.00608] [Medline: 32407171]
69. Mal PR, Suneel P, Shomeeta P. Social distancing: a non-pharmacological intervention for COVID-19. J Pak Med Assoc 2020 May;70(Suppl 3)(5):S21-S24. [doi: 10.5455/JPMA.05] [Medline: 32515371]
70. Nguyen TH. Lifting of social distancing measures: perspectives from Vietnam. Disaster Med Public Health Prep 2020 Jul 14:1-3 [FREE Full text] [doi: 10.1017/dmp.2020.238] [Medline: 32660690]
71. Tridandapani S, Holl G, Canon CL. Rapid deployment of home PACS workstations to enable social distancing in the coronavirus disease (COVID-19) era. American Journal of Roentgenology. Epub ahead of print posted online on May 20, 2020 2020. [doi: 10.2214/ajr.20.23495]

72. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. Int J Infect Dis 2020 May;94:91-95 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.017] [Medline: 32173574]
73. Chang T, Yang D, Wang M, Liang K, Tsai P, Chiou S, et al. Genomic analysis and comparative multiple sequences of SARS-CoV2. J Chin Med Assoc 2020 Jun;83(6):537-543. [doi: 10.1097/JCMA.0000000000000335] [Medline: 32349035]
74. Lehrer S, Rheinstein PH. SARS-CoV-2 gene sequence in the gene on human chromosome 1. In Vivo 2020 Jun;34(3 Suppl):1629-1632. [doi: 10.21873/invivo.11953] [Medline: 32503821]
75. Oude Munnink BB, Nieuwenhuijse DF, Stein M, O'Toole Á, Haverkate M, Mollers M, Dutch-Covid-19 response team. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. Nat Med. Epub ahead of print posted online on July 16, 2020 2020. [doi: 10.1038/s41591-020-0997-y] [Medline: 32678356]
76. Rockett RJ, Arnott A, Lam C, Sadsad R, Timms V, Gray K, et al. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. Nat Med. Epub ahead of print posted online on July 9, 2020 2020. [doi: 10.1038/s41591-020-1000-7] [Medline: 32647358]
77. Yap PSX, Tan TS, Chan YF, Tee KK, Kamarulzaman A, Teh CSJ. An overview of the genetic variations of the SARS-CoV-2 genomes isolated in Southeast Asian countries. J Microbiol Biotechnol 2020 Jul 28;30(7):962-966 [FREE Full text] [doi: 10.4014/jmb.2006.06009] [Medline: 32627759]
78. Chen S, Young M, Gounley J, Stanley C, Bhowmik D. Distinct structural flexibility within SARS-CoV-2 spike protein reveals potential therapeutic targets. bioRxiv. Preprint posted online on April 18, 2020. [doi: 10.1101/2020.04.17.047548]
79. Hao W, Ma B, Li Z, Wang X, Gao X, Li Y. Binding of the SARS-CoV-2 spike protein to glycans. bioRxiv. Preprint posted online on July 02, 2020 [FREE Full text] [doi: 10.1101/2020.05.17.100537]
80. Li Q, Peng W, Ou Y. Prediction and analysis of key protein structures of 2019-nCoV. Future Virology 2020 Jun;15(6):349-357. [doi: 10.2217/fvl-2020-0020]
81. Li T, Zheng Q, Yu H, Wu D, Xue W, Zhang Y. Characterization of the SARS-CoV-2 spike in an early prefusion conformation. bioRxiv. Preprint posted online on March 18, 2020. [doi: 10.1101/2020.03.16.994152]
82. Zhai X, Sun J, Yan Z, Zhang J, Zhao J, Zhao Z. Comparison of SARS-CoV-2 spike protein binding to human, pet, farm animals, and putative intermediate hosts ACE2 and ACE2 receptors. bioRxiv. Preprint posted online on May 08, 2020. [doi: 10.1101/2020.05.08.084061]
83. Banu N, Panikar SS, Leal LR, Leal AR. Protective role of ACE2 and its downregulation in SARS-CoV-2 infection leading to macrophage activation syndrome: therapeutic implications. Life Sci. Epub ahead of print posted online on June 03, 2020 2020 [FREE Full text] [doi: 10.1016/j.lfs.2020.117905]
84. Ratajczak MZ, Bujko K, Ciechanowicz A, Sielatycka K, Cymer M, Marlicz W, et al. SARS-CoV-2 entry receptor ACE2 is expressed on very small CD45 precursors of hematopoietic and endothelial cells and in response to virus spike protein activates the Nlrp3 inflammasome. Stem Cell Rev Rep. Epub ahead of print posted online on July 20, 2020 2020 [FREE Full text] [doi: 10.1007/s12015-020-10010-z]
85. Selçuk M, Çınar T, Keskin M, Çiçek V, Kılıç Ş, Kenan B, et al. Is the use of ACE inb/ARBs associated with higher in-hospital mortality in Covid-19 pneumonia patients? Clin Exp Hypertens. Epub ahead of print posted online on June 22, 2020 2020. [doi: 10.1080/10641963.2020.1783549]
86. South AM, Brady TM, Flynn JT. ACE2 (angiotensin-converting enzyme 2), COVID-19, and ACE inhibitor and Ang II (angiotensin II) receptor blocker use during the pandemic. Hypertension 2020 Jul;76(1):16-22. [doi: 10.1161/hypertensionaha.120.15291]
87. Verdecchia P, Cavallini C, Spanevello A, Angeli F. COVID-19: ACE2centric infective disease? Hypertension. Epub ahead of print posted online on June 1, 2020 2020. [doi: 10.1161/HYPERTENSIONAHA.120.15353] [Medline: 32476472]
88. Chan J, Sim Y, Ow S, Lim J, Kusumawidjaja G, Zhuang Q. The impact of COVID-19 on and recommendations for breast cancer care: the Singapore experience. Endocrine-Related Cancer 2020 Sep;27(9):307-327. [doi: 10.1530/erc-20-0157]
89. Gill S, Hao D, Hirte H, Campbell A, Colwell B. Impact of COVID-19 on Canadian medical oncologists and cancer care: Canadian Association of Medical Oncologists survey report. Curr Oncol 2020 Apr;27(2):71-74 [FREE Full text] [doi: 10.3747/co.27.6643] [Medline: 32489248]
90. Rodriguez J, Fletcher A, Heredia F, Fernandez R, Ramírez Salazar H, Sanabria D, et al. Alternative management for gynecological cancer care during the COVID-2019 pandemic: a Latin American survey. Int J Gynaecol Obstet. Epub ahead of print posted online on July 17, 2020 2020. [doi: 10.1002/ijgo.13272]
91. Werner MT, Carey RM, Albergotti WG, Lukens JN, Brody RM. Impact of the COVID-19 pandemic on the management of head and neck malignancies. Otolaryngol Head Neck Surg 2020 Jun;162(6):816-817. [doi: 10.1177/0194599820921413] [Medline: 32312163]
92. Sapkal G, Shete-Aich A, Jain R, Yadav PD, Sarkale P, Lakra R, $COVID Support Team: Pawar S., Patshute S., Salve V., Natarajan V., Marwah V., Malhotra B., Jain A., Sathe P., More R., Khedekar R., Suryawanshi D., Kalele K., Kumar A., Upadhyaya C.P., Waghmare A., Gawande P., Gopale S., Acharya M., Holeppanavar M., Thorat S., Chopade G., Vidhate S., Khutwad K., Phagiwala D., Bhanarkar S., Rakhe A., Salvi V., Dhaigude S., Ciyona. Development of indigenous IgG ELISA for the detection of anti-SARS-CoV-2 IgG. Indian J Med Res 2020 May;151(5):444-449 [FREE Full text] [doi: 10.4103/ijmr.IJMR_2232_20] [Medline: 32611915]

93. Schnurra C, Reiners N, Biemann R, Kaiser T, Trawinski H, Jassoy C. Comparison of the diagnostic sensitivity of SARS-CoV-2 nucleoprotein and glycoprotein-based antibody tests. J Clin Virol. Epub ahead of print posted online on July 6, 2020 2020 [FREE Full text] [doi: 10.1016/j.jcv.2020.104544] [Medline: 32663788]
94. Theel ES, Harring J, Hilgart H, Granger D. Performance characteristics of four high-throughput immunoassays for detection of IgG antibodies against SARS-CoV-2. J Clin Microbiol 2020 Jul 23;58(8):e01243-e01220 [FREE Full text] [doi: 10.1128/JCM.01243-20] [Medline: 32513859]
95. Tré-Hardy M, Wilmet A, Beukinga I, Favresse J, Dogné JM, Douxfils J, et al. Analytical and clinical validation of an ELISA for specific SARS-CoV-2 IgG, IgA, and IgM antibodies. J Med Virol. Epub ahead of print posted online on July 27, 2020 2020 [FREE Full text] [doi: 10.1002/jmv.26303] [Medline: 32667733]
96. Tuaillon E, Bolloré K, Pisoni A, Debiesse S, Renault C, Marie S, et al. Detection of SARS-CoV-2 antibodies using commercial assays and seroconversion patterns in hospitalized patients. J Infect. Epub ahead of print posted online on June 03, 2020 2020 [FREE Full text] [doi: 10.1016/j.jinf.2020.05.077] [Medline: 32504735]
97. Fischer RJ, Morris DH, van Doremalen N, Sarchette S, Matson MJ, Bushmaker T, et al. Effectiveness of N95 respirator decontamination and reuse against SARS-CoV-2 virus. Emerg Infect Dis. Epub ahead of print posted online on June 03, 2020 2020 [FREE Full text] [doi: 10.3201/eid2609.201524]
98. Rhee S. Management of used personal protective equipment and wastes related to COVID-19 in South Korea. Waste Manag Res. Epub ahead of print posted online on June 10, 2020 2020. [doi: 10.1177/0734242X20933343]
99. Střížová Z, Bartůňková J, Smrž D. Can wearing face masks in public affect transmission route and viral load in COVID-19? Cent Eur J Public Health 2020 Jun;28(2):161-162 [FREE Full text] [doi: 10.21101/cejph.a6290] [Medline: 32592550]
100. Temmesfeld MJ, Jakobsen RB, Grant P. Does a surgical helmet provide protection against aerosol transmitted disease? Acta Orthop. Epub ahead of print posted online on June 23, 2020 2020. [doi: 10.1080/17453674.2020.1771525]
101. Ceriello A, Standl E, Catrinoiu D, Itzhak B, Lalic NM, Rahelic D, et al. Issues for the management of people with diabetes and COVID-19 in ICU. Cardiovasc Diabetol 2020 Jul 20;19(1):114 [FREE Full text] [doi: 10.1186/s12933-020-01089-2] [Medline: 32690029]
102. Hartmann-Boyce J, Morris E, Goyder C, Kinton J, Perring J, Nunan D, et al. Diabetes and COVID-19: risks, management, and learnings from other national disasters. Diabetes Care. Epub ahead of print posted online on June 16, 2020 2020. [doi: 10.2337/dc20-1192] [Medline: 32546593]
103. Pfützner A, Lazzara M, Jantz J. Why do people with diabetes have a high risk for severe COVID-19 disease?-A dental hypothesis and possible prevention strategy. J Diabetes Sci Technol 2020 Jul;14(4):769-771 [FREE Full text] [doi: 10.1177/1932296820930287] [Medline: 32506937]
104. Shi Q, Zhang X, Jiang F, Zhang X, Hu N, Bimu C, et al. Clinical characteristics and risk factors for mortality of COVID-19 patients with diabetes in Wuhan, China: a two-center, retrospective study. Diabetes Care 2020 Jul;43(7):1382-1391. [doi: 10.2337/dc20-0598] [Medline: 32409504]
105. Indraccolo U. A pregnant woman and the SARS-CoV-2 infection: how are barriers easily crossed? Recenti Prog Med 2020 Apr;111(4):259-260. [doi: 10.1701/3347.33190] [Medline: 32319449]
106. Li J, Wang Y, Zeng Y, Song T, Pan X, Jia M, et al. Critically ill pregnant patient with COVID-19 and neonatal death within two hours of birth. Int J Gynaecol Obstet 2020 Jul;150(1):126-128. [doi: 10.1002/ijgo.13189] [Medline: 32369616]
107. Liu D, Li L, Wu X, Zheng D, Wang J, Yang L, et al. Pregnancy and perinatal outcomes of women with coronavirus disease (COVID-19) pneumonia: a preliminary analysis. AJR Am J Roentgenol 2020 Jul;215(1):127-132. [doi: 10.2214/AJR.20.23072] [Medline: 32186894]
108. Poon LC, Yang H, Kapur A, Melamed N, Dao B, Divakar H, et al. Global interim guidance on coronavirus disease 2019 (COVID-19) during pregnancy and puerperium from FIGO and allied partners: Information for healthcare professionals. Int J Gynaecol Obstet 2020 Jun;149(3):273-286. [doi: 10.1002/ijgo.13156] [Medline: 32248521]
109. Kulkarni P, Mahadevappa M, Alluri S. COVID-19 pandemic and the impact on the cardiovascular disease patient care. Curr Cardiol Rev 2020;16(3):173-177. [doi: 10.2174/1573403X16666200621154842] [Medline: 32564757]
110. Aigner C, Dittmer U, Kamler M, Collaud S, Taube C. COVID-19 in a lung transplant recipient. J Heart Lung Transplant 2020 Jun;39(6):610-611 [FREE Full text] [doi: 10.1016/j.healun.2020.04.004] [Medline: 32340870]
111. Akdur A, Karakaya E, Ayvazoglu Soy EH, Alshalabi O, Kirnap M, Arslan H, et al. Coronavirus disease (COVID-19) in kidney and liver transplant patients: a single-center experience. Exp Clin Transplant 2020 Jun;18(3):270-274 [FREE Full text] [doi: 10.6002/ect.2020.0193] [Medline: 32519617]
112. Husain SA, Dube G, Morris H, Fernandez H, Chang J, Paget K, et al. Early outcomes of outpatient management of kidney transplant recipients with coronavirus disease 2019. Clin J Am Soc Nephrol. Epub ahead of print posted online on May 18, 2020 2020. [doi: 10.2215/CJN.05170420]
113. Sharma V, Shaw A, Lowe M, Summers A, van Dellen D, Augustine T. The impact of the COVID-19 pandemic on renal transplantation in the UK. Clin Med (Lond) 2020 Jul;20(4):e82-e86 [FREE Full text] [doi: 10.7861/clinmed.2020-0183] [Medline: 32457133]
114. Journal of Medical Virology. Scopus.: Wiley-Blackwell; 2020. URL: https://www.scopus.com/sourceid/15968 [accessed 2020-08-20]

115. Scopus. International Journal of Environmental Research and Public Health. Scopus. URL: https://www.scopus.com/sourceid/144989 [accessed 2020-08-20]
116. Journal of Medical Virology. SciRev. URL: https://scirev.org/journal/journal-of-medical-virology/ [accessed 2020-08-13]
117. IJERPH Statistics: Overview. MDPI. URL: https://www.mdpi.com/journal/ijerph/stats [accessed 2020-08-13]
118. Abd-alrazaq A, Alajlani M, Alhuwail D, Schneider J, Al-Kuwari S, Shah Z, et al. Artificial Intelligence in the Fight Against COVID-19: Scoping Review. J Med Internet Res 2020 Dec 15;22(12) [FREE Full text] [doi: 10.2196/20756] [Medline: 33284779]
119. Chamola V, Hassija V, Gupta V, Guizani M. A comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, blockchain, and 5G in managing its impact. IEEE Access 2020;8:90225-90265. [doi: 10.1109/access.2020.2992341]
120. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
121. Kritikos M. Ten technologies to fight coronavirus: in depth analysis. European Parliamentary Research Service. 2020 Apr. URL: https://www.europarl.europa.eu/RegData/etudes/IDAN/2020/641543/EPRS_IDA(2020)641543_EN.pdf [accessed 2021-02-23]

Abbreviations

AI: artificial intelligence
CORD-19: COVID-19 Open Research Dataset
CSV: comma-separated values
JSON: JavaScript Object Notation
PCR: polymerase chain reaction
TF-IDF: term frequency–inverse document

Edited by G Eysenbach; submitted 27.08.20; peer-reviewed by M Alajlani, L Fernandez-Luque (NO AVAILABLE FOR REVIEWS); comments to author 01.10.20; revised version received 14.10.20; accepted 24.11.20; published 08.03.21.

Please cite as:
Abd-Alrazaq A, Schneider J, Mifsud B, Alam T, Househ M, Hamdi M, Shah Z
A Comprehensive Overview of the COVID-19 Literature: Machine Learning–Based Bibliometric Analysis
J Med Internet Res 2021;23(3):e23703
URL: https://www.jmir.org/2021/3/e23703
doi:10.2196/23703
PMID:33600346

Original Paper

Virtual Health Care for Community Management of Patients With COVID-19 in Australia: Observational Cohort Study

Owen Rhys Hutchings[1], MBBS; Cassandra Dearing[1], BN, LLB, MHSM; Dianna Jagers[1], BN, MHI; Miranda Jane Shaw[1], MHA; Freya Raffan[1], MHSM; Aaron Jones[2,3], MCE; Richard Taggart[2], MPharm; Tim Sinclair[2], DBA; Teresa Anderson[2], PhD; Angus Graham Ritchie[2,3,4], MBBS

[1]Royal Prince Alfred Virtual Hospital, Sydney Local Health District, Camperdown, Australia

[2]Sydney Local Health District, Camperdown, Australia

[3]Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[4]Menzies Centre for Health Policy, University of Sydney, Sydney, Australia

Corresponding Author:
Angus Graham Ritchie, MBBS
Sydney Local Health District
Missenden Road
Camperdown, 2050
Australia
Phone: 61 2 9767 6447
Email: angus.ritchie@health.nsw.gov.au

Abstract

Background: Australia has successfully controlled the COVID-19 pandemic. Similar to other high-income countries, Australia has extensively used telehealth services. Virtual health care, including telemedicine in combination with remote patient monitoring, has been implemented in certain settings as part of new models of care that are aimed at managing patients with COVID-19 outside the hospital setting.

Objective: This study aimed to describe the implementation of and early experience with virtual health care for community management of patients with COVID-19.

Methods: This observational cohort study was conducted with patients with COVID-19 who availed of a large Australian metropolitan health service with an established virtual health care program capable of monitoring patients remotely. We included patients with COVID-19 who received the health service, could self-isolate safely, did not require immediate admission to an in-patient setting, had no major active comorbid illness, and could be managed at home or at other suitable sites. Skin temperature, pulse rate, and blood oxygen saturation were remotely monitored. The primary outcome measures were care escalation rates, including emergency department presentation, and hospital admission.

Results: During March 11-29, 2020, a total of 162 of 173 (93.6%) patients with COVID-19 (median age 38 years, range 11-79 years), who were diagnosed locally, were enrolled in the virtual health care program. For 62 of 162 (38.3%) patients discharged during this period, the median length of stay was 8 (range 1-17) days. The peak of 100 prevalent patients equated to approximately 25 patients per registered nurse per shift. Patients were contacted a median of 16 (range 1-30) times during this period. Video consultations (n=1902, 66.3%) comprised most of the patient contacts, and 132 (81.5%) patients were monitored remotely. Care escalation rates were low, with an ambulance attendance rate of 3% (n=5), emergency department attendance rate of 2.5% (n=4), and hospital admission rate of 1.9% (n=3). No deaths were recorded.

Conclusions: Community-based virtual health care is safe for managing most patients with COVID-19 and can be rapidly implemented in an urban Australian setting for pandemic management. Health services implementing virtual health care should anticipate challenges associated with rapid technology deployments and provide adequate support to resolve them, including strategies to support the use of health information technologies among consumers.

(J Med Internet Res 2021;23(3):e21064) doi:10.2196/21064

KEYWORDS

COVID-19; digital health; health; informatics; remote monitoring; telehealth; virtual health care

Introduction

Australia has been remarkably successful in controlling the COVID-19 pandemic, having reported some of the lowest numbers of cases and deaths among other high-income countries [1]. This success is attributed to a rapid nationwide coordinated public health response to implement strong control measures that include widespread testing and contact tracing, social distancing, prohibition of public gatherings, use of face masks, and restrictions on international and domestic travel, including a mandatory quarantine period for those arriving from other countries [2,3]. The Australian government has also implemented nationwide funding for telehealth services, which is now a permanent reform, leading to rapid adoption of telehealth services [4].

COVID-19 manifests relatively mildly in most cases; however, approximately 14% of infected individuals develop severe disease that requires hospitalization, and 5% require admission to an intensive care unit [2]. The vast majority of COVID-19 cases in Australia have been actively monitored and managed by the public health system, and this arrangement was sustainable owing to the small number of cases and effective contact tracing systems. However, it was initially speculated that the Australian health care system had inadequate acute care facilities and intensive care unit beds to manage the high expected case load, which led to the rapid exploration of new models of acute care [5,6].

Virtual health care (VHC) is a model of health service delivery, which substitutes in-person consultations with telephonic or video consultations and often includes asynchronous data collection from the patient via survey tools with or without real-time remote monitoring. VHC is emerging as a central strategy to manage large numbers of medical patients affected by the COVID-19 pandemic, as this can maximize the use of limited clinical resources, reduce pressure on acute care facilities, and reduce the risk of health care–associated infections [7]. Studies have reported high satisfaction rates for telehealth services among patients and clinicians, with comparable clinical and service outcomes for chronic diseases [8]. Smith et al [9] reported on telehealth deployments in other crises including hurricanes Harvey and Irma. VHC-based models of acute care for COVID-19 management in high-income countries are emerging [10-17]. However, few models that include continuous remote monitoring of clinical observations and extensive utilization of videoconferencing platforms have been reported.

This study describes the rapid deployment of a VHC model that includes remote monitoring of clinical observations and routine use of videoconferencing platforms within a large Australian metropolitan public health service for managing outpatients with COVID-19.

Methods

Setting

Sydney Local Health District (SLHD) is a large metropolitan public health service in New South Wales, Australia, encompassing 5 hospitals, 4 large community health centers, and 12,000 staff. SLHD is responsible for the health and well-being of approximately 700,000 individuals living within its geographic boundaries and approximately 1,000,000 individuals who travel to the city each day for work, study, and recreation [18].

SLHD has been implementing telehealth and VHC modalities for several years. On February 3, 2020, SLHD commenced operations of the Royal Prince Alfred Virtual Hospital (rpavirtual), Australia's first metropolitan virtual hospital [19]. This service was established as a 12-month pilot program providing at-home and remote web-based nursing services. The initial patient cohorts included those seeking palliative care, adult patients with cystic fibrosis, and those at the risk of recurrent lower leg wounds. rpavirtual has a robust operational and clinical governing body including a general manager and a clinical director to oversee its operations, and it is embedded in the organizational structure of the health service, which includes a dedicated public health unit (PHU).

Colocated in the Royal Prince Alfred Hospital campus in Camperdown, New South Wales, rpavirtual is a 24/7 care center with technology-enabled multidisciplinary team rooms, handover areas, tracking boards, and several care pods. The care pods provide access to the electronic medical record (EMR) and shared care planning and remote monitoring tools. They are equipped with videoconferencing and telephone facilities and are staffed by nurses who can remotely monitor multiple patients simultaneously. The facility has medical staff on site and is under the supervision of a clinical director and director of nursing. A team of >100 community nurses are available to deliver in-home nursing care to complement VHC services.

On March 5, 2020, in response to the COVID-19 pandemic in Australia, rpavirtual began a rapid redesign of care systems to provide VHC to patients with COVID-19 managed in the community. The first patients were enrolled on March 11, 2020. This study describes the early experience with VHC use to manage patients with COVID-19 in the community up to and including March 29, 2020. The original patient cohorts continued receiving VHC and their outcomes are described separately.

The SLHD ethics review committee reviewed this study, and no ethical concerns were raised regarding the study or publication of the results.

Population

Patients who attended COVID-19 testing clinics of the SLHD and tested positive were informed of their outcomes by the local PHU and referred to the rpavirtual care center. The care center conducted an initial telephonic clinical assessment to ascertain suitability for VHC. Inclusion criteria and relative exclusion criteria are listed in Textbox 1.

Textbox 1. Selection criteria of the Royal Prince Alfred Virtual Hospital for virtual health care for patients with COVID-19.

Inclusion criteria

- The public health unit is satisfied that the patient can self-isolate safely and understands how to manage self-isolation.
- The clinical team is satisfied that it is clinically appropriate to manage the patient at home and a caregiver can safely provide care to the patient with appropriate personal protective equipment, or the patient can provide care to himself/herself, and a member of the clinical team carries out daily monitoring and follow-up evaluation.

Relative exclusion criteria

- Individuals over 65 years of age with significant comorbidity including, but not limited to, cancer, cardiovascular disease, diabetes, heart failure, immunosuppression, stroke, liver disease, renal disease, and lung disease.
- Individuals under 65 years of age with one or more of the following comorbidities: lung disease, cardiovascular disease, renal disease (including stage 5 chronic kidney disease or requiring renal replacement therapy including renal transplantation).
- Uncontrolled hypertension.
- Individuals at residential aged care facilities.
- Individuals aged <18 years.
- Pregnant women.

Patients matching one of the relative exclusion criteria were still accepted into VHC subject to further discussion with their treating physicians or an emergency medicine specialist. To be able to use the supplied technology and videoconferencing facilities, patients required access to a smartphone, tablet device, or personal computer with an internet connection and video capability.

Patients unable to self-isolate at home were offered alternative accommodation managed by the health service. Non–English-speaking patients were accepted with the support of interpreters. Patients deemed ineligible or unsuitable for VHC were required to be hospitalized. Multimedia Appendix 1 provides a schematic representation of the eligibility assessment process of rpavirtual for patients with COVID-19.

Model of Care

The rpavirtual COVID-19 VHC model was based on early detection of deterioration and managed care escalation for deteriorating patients.

Vital signs, including respiratory rate, oxygen saturation, pulse rate, and temperature, were monitored at home. Blood pressure monitoring was not required. The care center contacted patients at pre-arranged intervals thrice a day, including a video consultation with the patient twice every 24 hours, thus facilitating further assessment of symptoms and signs of deterioration (Textbox 2) based on standard nursing assessment approaches [20,21].

Textbox 2. Telemedicine assessment items of the Royal Prince Alfred Virtual Hospital for patients with COVID-19. aClinical assessment was conducted through video consultations, direct observation, patient-reported observations, or with remote monitoring devices (for temperature only).

Clinical progress

- Do you feel feverish or have chills?
- Do you have a cough?
- Do you feel short of breath or find it difficult to talk?
- Do you have any other symptoms?

Psychological screening

- How are you managing your home isolation?
- How are you coping generally?

Clinical assessment[a]

- Airway
- Breathing: respiratory distress, respiratory rate, oxygen saturation, and other respiratory symptoms
- Cardiovascular system: appearance and heart rate
- Disability: alertness, cognition, mental state, and mobility
- Exposure (temperature): temperature
- Fluid balance: fluid and food intake and gastrointestinal symptoms and losses
- Glycemic control (if diabetic): blood glucose levels

Vital signs were recorded electronically in the EMR and tracked against a standardized early warning system known as "Between The Flags," using the standard criteria for observing adults in the general population (Multimedia Appendix 2) [22,23].

Clinical escalation of deteriorating patients was carried out through managed transfer to the local emergency department in an ambulance if clinically indicated. The ambulance service was notified of the patient's infectious status and advised to contact the receiving emergency department prior to arrival. In the event of escalation, the care center also notified the hospital executive and the local PHU.

An escalation communication pathway was activated if a patient could not be contacted by the care center. If a patient could not be contacted after one telephone call, a text message was sent. If the patient did not revert within 1 hour, a second text message was sent, which expressed concern for their welfare. If the patient did not respond to a third telephone call, the New South Wales Police was contacted to conduct a welfare check.

Medical officers at rpavirtual were consulted through referral from the care center staff to discuss patient deterioration, escalation decision-making, medical certification, and prescribing medications, although medication management was not a component of the model of care.

Discharge from VHC was managed in accordance with the following "release from isolation" criteria [24]: the individual has been afebrile for the previous 48 hours, the acute illness resolved in the previous 24 hours, at least 7 days have elapsed since the onset of the acute illness, and a negative result was obtained on RT–PCR with at least two consecutive respiratory specimens collected 24 hours apart after the acute illness has resolved.

Patients meeting the first three criteria were referred to a COVID-19 testing clinic for repeat testing. On obtaining negative results, the patient was discharged from VHC and referred to their general practitioner for ongoing care.

Patient Experience

The use of technology by patients and their caregivers is central to care delivery by rpavirtual. Once enrolled in VHC, new patients were provided with a welcome pack containing the following items: a welcome letter and videoconferencing instructions, pulse oximeter and instructions for its use, a temperature monitoring device and instructions for its use, New South Wales Government COVID-19 factsheets, and a patient responsibilities pamphlet.

Personal protective equipment (PPE) was provided immediately after COVID-19 testing and then resupplied with the welcome pack.

The welcome pack was delivered by a "Flying Squad" of health informatics staff wearing appropriate PPE upon completion of a home visit risk assessment. The rpavirtual Flying Squad instructed patients to wear PPE when answering the door and notified patients when they were 5 minutes away from their homes.

Patients interacted with the rpavirtual team through video consultations and a tollfree telephone service that operated 24/7. Calls were scheduled with the patient to collect their self-observations. Patients were advised to call the rpavirtual call center or an ambulance if they experienced deterioration.

Remote Monitoring Technology

Remote monitoring devices were evaluated to collect patient-generated health data from patients with COVID-19 in accordance with defined criteria (Textbox 3).

Textbox 3. Criteria of the Royal Prince Alfred Virtual Hospital for remote patient monitoring technology (in random order).

- Capable of measuring the desired clinical parameters (ie, heart rate, oxygen saturation, temperature, and respiratory rate)
- Capable of remote data collection and transmission (or able to be read by the patient)
- Usability for the medical team at Royal Prince Alfred Virtual Hospital
- Usability for patients with COVID-19 admitted to the Royal Prince Alfred Virtual Hospital
- Training and support requirements for the health informatics staff of Sydney Local Health District
- Listed on the Australian Register of Therapeutic Goods of the Therapeutic Goods Administration (Department of Health, Australian Government) or able to be fast-tracked (eg, Conformité Européenne– or US Food and Drug Administration–marked)
- Supply chain availability to meet the scale, speed, and demand of the COVID-19 response
- Cost
- Compliance with cyber security, data security, and privacy and infection control requirements

No suitable single device that met all vital sign monitoring requirements in accordance with the aforementioned criteria was identified. Two devices were selected for use in the COVID-19 program: a pulse oximeter and temperature patch (Figure 1). No reliable device for measuring the respiratory rate was identified; respiratory rate was measured through videoconference consultations.

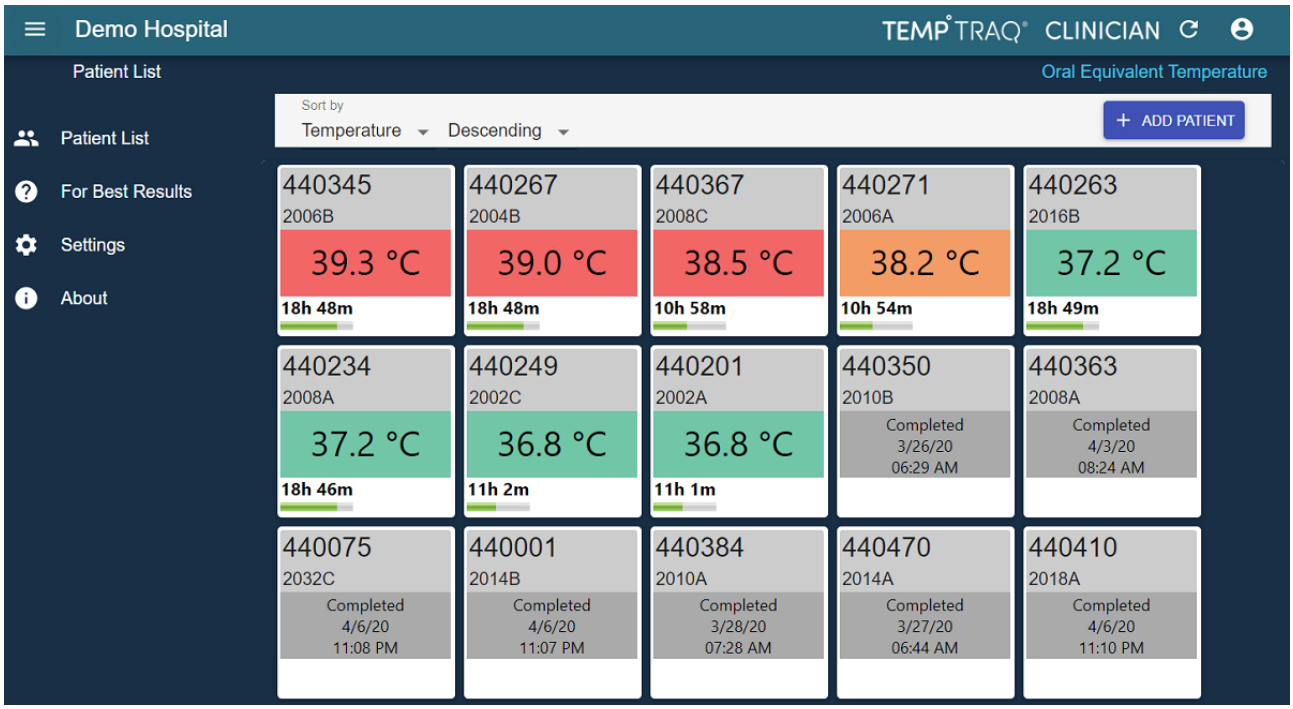
Figure 1. Remote monitoring equipment provided to patients for measurement of pulse, blood oxygen saturation, and temperature.

A wireless pulse oximeter (iHealth Air pulse oximeter PO3M, iHealth Labs, Inc.) facilitated peripheral oxygen saturation and pulse rate measurements. Pulse oximeters were only used by one patient; they were not reused as they could not be adequately disinfected. A single-use, wearable temperature monitor (Temp°Traq Clinical, Blue Spark Technologies, Inc.) was self-applied in the axilla and facilitated continuous temperature monitoring. The device fed data continuously into a web-based dashboard, providing the care center with an overview of all patients. Each patch lasted 72 hours, and each patient was provided with 3 patches to cover the first 9-11 days of isolation.

Both devices had Bluetooth connectivity, but only the temperature monitor required connection with a compatible Apple or Android smartphone to be read; readings from the pulse oximeter were read directly from the device and reported through video consultations. Both devices were approved by the US Food and Drug Administration and Conformité Européenne–marked and either registered with the Therapeutic Goods Administration (Department of Health, Australian Government) or able to be fast-tracked but had only limited prior evaluation in the clinical setting [25,26].

The Flying Squad prepared devices for delivery, including precharging of the pulse oximeters and registering of all 3 temperature monitoring patches against the specific patient in the monitoring portal. After delivery, they telephonically contacted the patient to help them download and set up the required mobile app, allowing temperature readings to be fed continuously into a web-based dashboard. The care center had access to an overview of all prevalent monitored patients with color-coded parameters to identify patients with an abnormal temperature or patches that did not function normally (Figure 2).

Figure 2. Screenshot of the temperature monitoring portal. This figure is from a demonstration system, and the temperature color indicators shown here are configured differently from the version in use: ≥38.0 C=red, 37.5 C-37.9 C=orange, 36.0 C-37.4 C=green, <36.0 C=blue. Expired or nonfunctioning patches are indicated in grey.



Patient-reported observations were manually documented in the EMR (Cerner Millennium, Cerner Corp) in purpose-built sections to allow them to be differentiated from other sources of vital signs.

Video Consultations

Video consultations were carried out to visually assess the patient, confirm vital signs recorded by wearable devices, and to estimate the respiratory rate.

The rpavirtual care pods were fitted with high-definition webcams and Bluetooth-enabled headsets to allow the patient to see and hear the nurse. The care pods were configured to facilitate privacy during conversations with the patients. Pull-up backgrounds and feature walls were present behind each nurse to minimize distractions on video.

A commercial web-based video conferencing solution Pexip (Pexip AS) was used for remote video consultations. This platform functioned on both desktop and mobile devices and was endorsed by eHealth NSW, the state government agency overseeing the use of technology in health care. Patients called in to video consultations using their own computer or mobile device.

Results

In March 11-29, 2020, 5821 individuals were tested for COVID-19 at SLHD COVID-19 testing clinics, and 173 individuals tested positive. Of them, 162 (93.6%) were enrolled in the rpavirtual program, and the remaining individuals were admitted to hospital or referred to another health service. The median age of the admitted patients was 38 (range 11-79) years; 3 patients aged <18 years and 2 pregnant women were admitted to VHC.

The median number of admissions per day was 6 (range 1-30). A total of 62 (38.3%) patients were discharged during this period, with a median length of stay of 8 (range 1-17) days. At the end of this period, 100 prevalent patients received VHC.

The care center commenced operations with 4 full-time–equivalent registered nurses and gradually increased to 9.5 full-time–equivalent at the end of the period, which

equated to a ratio of approximately 25 patients per registered nurse per shift.

Patients were contacted 2865 times, with a median of 16 (range 1-50) contacts per patient. Video consultations (n=1902, 66.3%) comprised the majority of patient contact, and telephonic consultations (n=688, 24.0%) accounted for the remainder of patient contact. The ratio of telephonic-to-video consultations was 1:2.8. The median duration of each contact was 8.5 (IQR 5-15) minutes for telephonic consultations and 15 (IQR 13-15) minutes for video consultations.

During March 18-29, 2020, the rpavirtual Flying Squad delivered welcome packs with remote monitoring equipment to 132 of 162 (81.5%) patients.

Ambulances were called to attend to 5 patients, and 4 patients were transferred to the emergency department for assessment. In total, 3 patients were subsequently admitted, and 1 was discharged to continue with VHC at home. No deaths or police welfare checks were recorded. However, one individual was subject to a public health order for failing to adhere to self-isolation requirements. Detailed patient outcomes will be reported separately.

Discussion

Principal Findings

This study describes the rapid implementation of a model of care and technology to deliver VHC for community management of patients with COVID-19. We found that by excluding high-risk patients with COVID-19, we could include most individuals testing positive upon local diagnosis. Through remote clinical appraisal, supported by remote monitoring of clinical observations, we observed low rates of deterioration, with few patients requiring clinical escalation and no patient deaths. The program also enrolled pregnant women and pediatric patients, expanding the range of patients who may be suitable for VHC for COVID-19 as an alternative to hospital in-patient admission.

A range of technical and operational issues were expected with the rapid implementation of video consultations [8]. Our model of care required 3 patient contacts per day, with at least 2 contacts on video, to ensure adequate surveillance of signs of deterioration and to measure the respiratory rate directly. Initially, patients were scheduled for 10-minute video and telehealth consultations. On encountering a problem with video consultation, which the staff could not rapidly resolve, a telephonic consultation was carried out instead. This is an example of workaround—a common phenomenon when using health information technology that is driven by the need to resolve conflicting goals in a timely manner [27]. In this case, the workaround was driven by the need to adhere to the schedule of appointments rather than delaying to resolve the issue. Improving the user experience for staff and patients required both technical and workflow changes, and fixed appointment slots were abandoned. The optimal approach to training and support in VHC to avoid workarounds, particularly with a rapid increase in staff redeployed from other areas of the business (in the context of high service demands) requires further consideration.

The need for enhanced support for patients to use health information technologies was identified early in the implementation. The model of care relied on the ability of patients or their caregivers to use medical technology (pulse oximeter and temperature patch) and digital health applications (for temperature monitoring and video consultations). The use of health information technology by consumers, known as consumer health informatics, has a strong focus on usability and accessibility [28,29]. We immediately realized that patients enrolled in VHC could not easily download the temperature monitoring application and connect it with the patches with the instructions provided in the welcome pack alone. The welcome packs delivered by the Flying Squad, a multidisciplinary health informatics team, were ideally suited for providing additional support. A process was established to contact patients shortly after delivery of the welcome pack and support them in achieving the goal of transmitting temperature readings to the cloud-based monitoring panel. If consumer health technology is to be relied on for health care delivery, health services will require strategies to support it. While this has been explored in the chronic disease setting, in which patients can avail of in-person health services, this is not feasible during an infectious disease pandemic [30]. Managed mobile health care platforms, in which patients are provided with a tablet computer with all relevant applications pre-installed and connected to relevant peripherals, may simplify and improve the patient experience, and rpavirtual has used this approach with other patient cohorts. The role of health informaticians in supporting consumer health informatics has received limited attention and warrants further exploration.

A uniform model of care for all patients, regardless of care needs or the risk of deterioration, may not be appropriate or necessary. Risk stratification upon admission with enhanced monitoring of patients at a higher risk of deterioration was considered, although at the time, limited empirical evidence was available to guide that strategy [31]. Pulse oximetry appears to be a useful tool for risk stratification. Patients with COVID-19 experience deterioration typically on day 7 of symptom onset, which was the median period of hospitalization for patients who developed an associated pneumonia [32]. Patient-reported pulse oximetry measurements are effective for detecting silent hypoxia and predicting hospitalization; this supports the use of this technology for patient monitoring and risk stratification [16]. In addition to the primary management of COVID-19, pulse oximetry has effectively facilitated decisions on the early discharge of patients from hospital for a cohort of patients with severe COVID-19, with particular benefits among patients with a persistent need for oxygen therapy [13]. A risk-stratified care pathway has now been developed and is being used to guide ongoing patient management (Multimedia Appendix 3).

Rapid changes to the EMR are required to support VHC for patients with COVID-19. EMRs are potentially useful tools for rapid deployment of standardized processes, including responses to the COVID-19 pandemic [33]. To support the redesign of rpavirtual to provide care to patients with COVID-19, new locations and patient lists were generated in the EMR. Access

to the electronic Between the Flags record for detecting signs of deterioration, previously only used in the in-patient setting, was extended to all care center staff for community use. Clinical documentation templates, simple reports, and modifications to the results flowsheet to facilitate recording and clinical review of patient-reported and remotely monitored clinical observations were designed and implemented. However, communication systems were managed separately from the EMR, with separate systems used for text messaging, telephone calls, and videoconferencing. A patient portal would have aided communication and recording of patient-reported measures, as previously reported [14,34]. A comprehensive and integrated suite of digital health care tools would reduce fragmentation of information systems and workflows and improve service delivery by automating manual processes.

Mental health and well-being issues require further consideration in the provision of VHC to patients with COVID-19 in Australia owing to extensive quarantine and self-isolation periods. Such measures for controlling infectious diseases have been associated with negative psychological effects including posttraumatic stress symptoms, confusion, and anger, with the potential to be long-lasting [35]. Two questions directed toward psychological assessment were included in the model to help recognize mental health concerns. The psychological impact of self-isolation in one's own home may be different from that of individuals placed in mandatory quarantine in hotels after arriving from other countries, which has been a key strategy in Australia [3]. This issue requires further study with consideration to provide access to social workers, psychologists, and psychiatrists through VHC. The key learnings from our experience are summarized in Textbox 4.

Textbox 4. Key learnings from the use of virtual health care for acute management of patients with COVID-19 in Australia.

- The acute nature of COVID-19 and the potential for rapid patient deterioration required 24/7 operations.
- A ratio of approximately 1 nurse per 25 patients per shift was required to support a model of care with high extensive of videoconferencing and continuous monitoring of patient observations with 24/7 operations; however, this could be reduced during low-activity periods such as during nighttime when patient contact was not scheduled.
- There was a high administrative burden in managing communication and interactions with enrolled patients when virtual health care is used without a digital patient engagement platform, such as a patient portal or other customer relationship management system with an integrated communications suite.
- Pulse oximetry appears to be an important tool in virtual care models for COVID-19 management for risk stratification and detection of deterioration and is feasible to use in the community. Readings that can be directly obtained from the device are useful.
- Consumers should be provided education, training, and support to use health technology if it is to be relied on for monitoring and management in virtual care settings.
- Social isolation measures introduce mental health risks that need to be considered within models of care, including screening tools and virtual access to appropriate services.

Comparison With Previous Studies

The use of VHC for managing outpatients with COVID-19 has been widely reported with small cohorts, at single centers, or in health systems. A consistent finding is that VHC is safe and effective for hospital avoidance, with low rates of clinical care escalation to the emergency department or other in-person observations. However, there is significant heterogeneity in the models of care and technology models, with few reports on programs that highly utilize video consultations combined with remote patient monitoring of clinical observations, including pulse oximetry and temperature, along with 24/7 support.

The only other study from Australia, which described the generation and use of a virtual ward within an existing hospital, involved the referral of patients by the local public health service, and patients were contacted once or twice a day only by telephone [17].

Studies on the use of VHC for COVID-19 management in the United States have reported the extensive use of patient portals, patient-reported symptoms and observations, and the use of telephonic consultations, with a reduced use of remote patient monitoring systems and videoconferencing. One large health system used symptom questionnaires on a daily basis to stratify and prioritize enrolled patients with suspected or confirmed COVID-19 for telephonic consultations. This system reported low rates of escalation to the emergency department; however, there was no access to remote monitoring or video consultations. With increasing experience, a triage system was adopted to exclude low-risk patients owing to the low incidence of deterioration [14]. Another institution reported the need to expand their workforce to provide a 24-hour service because they found that the enrolled patients were sending text messages to their remote patient monitoring app after the service was discontinued each day at 5 PM [12].

Expansion of a model of VHC to the in-patient setting, based on continuous video observation, has also been described. Patients admitted for COVID-19 management were placed in negative pressure rooms equipped with video monitoring systems to facilitate 2-way video and audio communication with ward health care staff [34]. This reduced the exposure of health care workers to COVID-19 and the consumption of PPE.

Limitations

This study has several limitations. First, this study has a small cohort, relative to other studies, and describes short-term experiences. Second, the technology used in this study was selected on the basis of pragmatic considerations, limiting opportunity for more thorough evaluation prior to use in practice. Finally, this study describes early results from a single health system in Australia and may not be generalizable to other

locations. Australia has seen relatively smaller COVID-19 caseloads than other high-income countries, implying that the demand for health services did not exceed their current capacity to the same extent as that in other countries. Significant heterogeneity has been in described virtual health models of care for COVID-19, which are strongly influenced by context, pragmatism, and local constraints. Even so, sharing our experience may help inform others as they continue dealing with the pandemic.

Conclusions

In summary, community-based VHC is a feasible and safe approach for managing less severe cases of COVID-19 and can be rapidly implemented in the Australian context for pandemic management with strong operational and clinical governance, including integration with clinical specialists. Health services implementing VHC should anticipate challenges with rapid technology implementations and provide adequate support to resolve them, including strategies to support consumer use of health information technologies.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Care centre eligibility criteria decision tree for COVID-19 management by rpavirtual.
[PDF File (Adobe PDF File), 529 KB - jmir_v23i3e21064_app1.pdf]

Multimedia Appendix 2
Printout of 48h view of Standard Adult General Observation Chart with recorded clinical observations as shown in the electronic medical record.
[PDF File (Adobe PDF File), 272 KB - jmir_v23i3e21064_app2.pdf]

Multimedia Appendix 3
rpavirtual COVID-19 risk stratification and clinical pathways.
[PDF File (Adobe PDF File), 113 KB - jmir_v23i3e21064_app3.pdf]

References

1. Australia. World Health Organization. URL: https://covid19.who.int/region/wpro/country/au [accessed 2021-02-19]
2. Chang SL, Harding N, Zachreson C, Cliff OM, Prokopenko M. Modelling transmission and control of the COVID-19 pandemic in Australia. Nat Commun 2020 Nov 11;11(1):5710 [FREE Full text] [doi: 10.1038/s41467-020-19393-6] [Medline: 33177507]
3. Fotheringham P, Anderson T, Shaw M, Jewitt J, Storey H, Hutchings O, et al. Control of COVID-19 in Australia through quarantine: the role of special health accommodation (SHA) in New South Wales, Australia. BMC Public Health 2021 Jan 27;21(1):225 [FREE Full text] [doi: 10.1186/s12889-021-10244-7] [Medline: 33504347]
4. Thomas EE, Haydon HM, Mehrotra A, Caffery LJ, Snoswell CL, Banbury A, et al. Building on the momentum: Sustaining telehealth beyond COVID-19. J Telemed Telecare 2020 Sep 26:1357633X20960638. [doi: 10.1177/1357633X20960638] [Medline: 32985380]
5. Fox G, Trauer J, McBryde E. Modelling the impact of COVID-19 on intensive care services in New South Wales. Med J Aust 2020 Jun;212(10):468-469 [FREE Full text] [doi: 10.5694/mja2.50606] [Medline: 32383153]
6. Litton E, Bucci T, Chavan S, Ho Y, Holley A, Howard G, et al. Surge capacity of intensive care units in case of acute increase in demand caused by COVID-19 in Australia. Med J Aust 2020 Jun 13;212(10):463-467 [FREE Full text] [doi: 10.5694/mja2.50596] [Medline: 32306408]
7. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
8. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
9. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: Implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Jun;26(5):309-313 [FREE Full text] [doi: 10.1177/1357633X20916567] [Medline: 32196391]
10. Thornton J. The "virtual wards" supporting patients with covid-19 in the community. BMJ 2020 Jun 04;369:m2119. [doi: 10.1136/bmj.m2119] [Medline: 32499317]
11. Torjesen I. Covid-19: Patients to use pulse oximetry at home to spot deterioration. BMJ 2020 Oct 27;371:m4151. [doi: 10.1136/bmj.m4151]

12. Annis T, Pleasants S, Hultman G, Lindemann E, Thompson J, Billecke S, et al. Rapid implementation of a COVID-19 remote patient monitoring program. J Am Med Inform Assoc 2020 Aug 01;27(8):1326-1330 [FREE Full text] [doi: 10.1093/jamia/ocaa097] [Medline: 32392280]
13. Grutters L, Majoor K, Mattern E, Hardeman J, van Swol CFP, Vorselaars A. Home telemonitoring makes early hospital discharge of COVID-19 patients possible. J Am Med Inform Assoc 2020 Nov 01;27(11):1825-1827 [FREE Full text] [doi: 10.1093/jamia/ocaa168] [Medline: 32667985]
14. Kricke G, Roemer P, Barnard C, Peipert J, Henschen B, Bierman J. Rapid Implementation of an Outpatient Covid-19 Monitoring Program. NEJM Catalyst Innovations in Care Delivery 2020 [FREE Full text] [doi: 10.1056/CAT.20.0214]
15. Lam PW, Sehgal P, Andany N, Mubareka S, Simor AE, Ozaldin O, et al. A virtual care program for outpatients diagnosed with COVID-19: a feasibility study. CMAJ Open 2020;8(2):E407-E413 [FREE Full text] [doi: 10.9778/cmajo.20200069] [Medline: 32447283]
16. Shah S, Majmudar K, Stein A, Gupta N, Suppes S, Karamanis M, et al. Novel Use of Home Pulse Oximetry Monitoring in COVID-19 Patients Discharged From the Emergency Department Identifies Need for Hospitalization. Acad Emerg Med 2020 Aug;27(8):681-692 [FREE Full text] [doi: 10.1111/acem.14053] [Medline: 32779828]
17. Ferry OR, Moloney EC, Spratt OT, Whiting GFM, Bennett CJ. A Virtual Ward Model of Care for Patients With COVID-19: Retrospective Single-Center Clinical Study. J Med Internet Res 2021 Feb 10;23(2):e25518 [FREE Full text] [doi: 10.2196/25518] [Medline: 33529157]
18. Year in Review 2018-19. Sydney Local Health District. 2019. URL: https://www.slhd.nsw.gov.au/pdfs/YIR_18-19.pdf [accessed 2020-04-05]
19. McDonald K. Sydney LHD trialling virtual hospital for community care from RPA. Pulse+IT. 2020 Mar 03. URL: https://www.pulseitmagazine.com.au/news/australian-ehealth/5364-sydney-lhd-trialling-virtual-hospital-for-community-care-from-rpa [accessed 2020-03-28]
20. Cathala X, Moorley C. Performing an A-G patient assessment: a practical step-by-step guide. Nursing Times. 2020 Jan 06. URL: https://www.nursingtimes.net/clinical-archive/assessment-skills/performing-an-a-g-patient-assessment-a-step-by-step-guide-06-01-2020/ [accessed 2021-03-03]
21. Smith D, Bowden T. Using the ABCDE approach to assess the deteriorating patient. Nursing Standard. 2017. URL: https://journals.rcni.com/nursing-standard/using-the-abcde-approach-to-assess-the-deteriorating-patient-ns.2017.e11030 [accessed 2021-03-04]
22. Between the Flags 2020. New South Wales Government: Clinical Excellence Commission. URL: http://cec.health.nsw.gov.au/keep-patients-safe/Deteriorating-patients/between-the-flags [accessed 2020-03-30]
23. Pain C, Green M, Duff C, Hyland D, Pantle A, Fitzpatrick K, et al. Between the flags: implementing a safety-net system at scale to recognise and manage deteriorating patients in the New South Wales Public Health System. Int J Qual Health Care 2017 Feb 01;29(1):130-136. [doi: 10.1093/intqhc/mzw132] [Medline: 27920243]
24. Coronavirus Disease 2019 (COVID-19). Australian Government: Department of Health. URL: https://www1.health.gov.au/internet/main/publishing.nsf/Content/cdna-song-novel-coronavirus.htm [accessed 2020-03-30]
25. Moreno-Alsasua L, Garcia-Zapirain B, David Rodrigo-Carbonero J, Ruiz IO, Hamrioui S, de la Torre Díez I. Primary Prevention of Asymptomatic Cardiovascular Disease Using Physiological Sensors Connected to an iOS App. J Med Syst 2017 Oct 26;41(12):191. [doi: 10.1007/s10916-017-0840-2] [Medline: 29075920]
26. Sampson M, Hickey V, Huber J, Alonso PB, Davies SM, Dandoy CE. Feasibility of continuous temperature monitoring in pediatric immunocompromised patients: A pilot study. Pediatr Blood Cancer 2019 Jun;66(6):e27723. [doi: 10.1002/pbc.27723] [Medline: 30884117]
27. Patterson ES. Workarounds to Intended Use of Health Information Technology: A Narrative Review of the Human Factors Engineering Literature. Hum Factors 2018 May;60(3):281-292. [doi: 10.1177/0018720818762546] [Medline: 29533682]
28. Eysenbach G. Consumer health informatics. BMJ 2000 Jun 24;320(7251):1713-1716 [FREE Full text] [doi: 10.1136/bmj.320.7251.1713] [Medline: 10864552]
29. Goldberg L, Lide B, Lowry S, Massett HA, O'Connell T, Preece J, et al. Usability and accessibility in consumer health informatics current trends and future challenges. Am J Prev Med 2011 May;40(5 Suppl 2):S187-S197. [doi: 10.1016/j.amepre.2011.01.009] [Medline: 21521594]
30. Milani RV, Bober RM, Lavie CJ. The Role of Technology in Chronic Disease Care. Prog Cardiovasc Dis 2016;58(6):579-583. [doi: 10.1016/j.pcad.2016.01.001] [Medline: 26772623]
31. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A Tool for Early Prediction of Severe Coronavirus Disease 2019 (COVID-19): A Multicenter Study Using the Risk Nomogram in Wuhan and Guangdong, China. Clin Infect Dis 2020 Jul 28;71(15):833-840 [FREE Full text] [doi: 10.1093/cid/ciaa443] [Medline: 32296824]
32. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
33. Reeves J, Hollandsworth H, Torriani F, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. J Am Med Inform Assoc 2020 Jun 01;27(6):853-859 [FREE Full text] [doi: 10.1093/jamia/ocaa037] [Medline: 32208481]

34. Ford D, Harvey J, McElligott J, King K, Simpson K, Valenta S, et al. Leveraging health system telehealth and informatics infrastructure to create a continuum of services for COVID-19 screening, testing, and treatment. J Am Med Inform Assoc 2020 Dec 09;27(12):1871-1877 [FREE Full text] [doi: 10.1093/jamia/ocaa157] [Medline: 32602884]
35. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]

Abbreviations

EMR: electronic medical record
PPE: personal protective equipment
rpavirtual: Royal Prince Alfred Virtual Hospital
SLHD: Sydney Local Health District
VHC: virtual health care

Edited by G Eysenbach, C Basch; submitted 05.06.20; peer-reviewed by A Lau, F Muhammad, I Mircheva, K Goniewicz, M Adly, A Adly, A Adly, MDG Pimentel; comments to author 28.10.20; revised version received 17.11.20; accepted 01.03.21; published 09.03.21.

Please cite as:
Hutchings OR, Dearing C, Jagers D, Shaw MJ, Raffan F, Jones A, Taggart R, Sinclair T, Anderson T, Ritchie AG
Virtual Health Care for Community Management of Patients With COVID-19 in Australia: Observational Cohort Study
J Med Internet Res 2021;23(3):e21064
URL: https://www.jmir.org/2021/3/e21064
doi:10.2196/21064
PMID:33687341

Original Paper

Behavioral Intention to Receive a COVID-19 Vaccination Among Chinese Factory Workers: Cross-sectional Online Survey

Ke Chun Zhang[1], MSc; Yuan Fang[2], PhD; He Cao[1], MSc; Hongbiao Chen[1], MSc; Tian Hu[1], BSc; Yaqi Chen[1], BSc; Xiaofeng Zhou[1], BSc; Zixin Wang[3], PhD

[1]Longhua District Center for Disease Control and Prevention, Shenzhen, China

[2]Department of Early Childhood Education, The Education University of Hong Kong, Hong Kong, China

[3]JC School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, China

Corresponding Author:
Zixin Wang, PhD
JC School of Public Health and Primary Care
The Chinese University of Hong Kong
Room 508, School of Public Health, Prince of Wales Hospital, Shatin, N.T.
Hong Kong, 666888
China
Phone: 852 22528740
Email: wangzx@cuhk.edu.hk

Abstract

Background: COVID-19 vaccines will become available in China soon. Understanding communities' responses to the forthcoming COVID-19 vaccines is important. We applied the theory of planned behavior as the theoretical framework.

Objective: This study investigates the prevalence of and factors associated with behavioral intention to receive self-financed or free COVID-19 vaccinations among Chinese factory workers who resumed work during the pandemic. We examined the effects of factors including sociodemographics, perceptions related to COVID-19 vaccination, exposure to information about COVID-19 vaccination through social media, and COVID-19 preventive measures implemented by individuals and factories.

Methods: Participants were full-time employees 18 years or older who worked in factories in Shenzhen. Factory workers in Shenzhen are required to receive a physical examination annually. Eligible workers attending six physical examination sites were invited to complete a survey on September 1-7, 2020. Out of 2653 eligible factory workers, 2053 (77.4%) completed the online survey. Multivariate two-level logistic regression models and ordinal logistic regression models were fitted.

Results: The prevalence of behavioral intention to receive a COVID-19 vaccination was 66.6% (n=1368, conditional on 80% vaccine efficacy and market rate) and 80.6% (n=1655, conditional on 80% vaccine efficacy and free vaccines). After adjusting for significant background characteristics, positive attitudes toward COVID-19 vaccination (adjusted odds ratio [AOR] 1.20, 95% CI 1.15-1.25 and AOR 1.24, 95% CI 1.19-1.30), perceived support from significant others for getting a COVID-19 vaccination (AOR 1.43, 95% CI 1.32-1.55 and AOR 1.37, 95% CI 1.25-1.50), and perceived behavioral control to get a COVID-19 vaccination (AOR 1.51, 95% CI 1.32-1.73 and AOR 1.28, 95% CI 1.09-1.51) were positively associated with both dependent variables (conditional on 80% vaccine efficacy and market rate or free vaccines, respectively). Regarding social media influence, higher frequency of exposure to positive information related to COVID-19 vaccination was associated with a higher intention to receive a COVID-19 vaccination at market rate (AOR 1.53, 95% CI 1.39-1.70) or a free vaccination (AOR 1.52, 95% CI 1.35-1.71). Higher self-reported compliance with wearing a face mask in the workplace (AOR 1.27, 95% CI 1.02-1.58 and AOR 1.67, 95% CI 1.24-2.27) and other public spaces (AOR 1.80, 95% CI 1.42-2.29 and AOR 1.34, 95% CI 1.01-1.77), hand hygiene (AOR 1.21, 95% CI 1.00-1.47 and AOR 1.52, 95% CI 1.19-1.93), and avoiding social gatherings (AOR 1.22, 95% CI 1.01-1.47 and AOR 1.55, 95% CI 1.23-1.95) and crowded places (AOR 1.24, 95% CI 1.02-1.51 and AOR 1.73, 95% CI 1.37-2.18) were also positively associated with both dependent variables. The number of COVID-19 preventive measures implemented by the factory was positively associated with the intention to receive a COVID-19 vaccination under both scenarios (AOR 1.08, 95% CI 1.04-1.12 and AOR 1.06, 95% CI 1.01-1.11).

Conclusions: Factory workers in China reported a high behavioral intention to receive a COVID-19 vaccination. The theory of planned behavior is a useful framework to guide the development of future campaigns promoting COVID-19 vaccination.

(J Med Internet Res 2021;23(3):e24673) doi:10.2196/24673

KEYWORDS

COVID-19; vaccination; behavioral intention; perception; social media influence; personal preventive behaviors; factory workers; China; social media; vaccine; behavior; intention; risk

Introduction

The COVID-19 pandemic remains out of control worldwide. As of September 20, 2020, there were 30,675,675 confirmed cases and 954,417 deaths [1]. Since immunization against COVID-19 is not yet available, the current means for pandemic control is to avoid exposure. These measures (eg, physical distancing and lockdown) are likely detrimental to the global economy. Except for China, all G20 countries, a group of the world's largest economies, experienced a decrease in gross domestic product in the second quarter of 2020 due to the COVID-19 pandemic [2]. Moreover, implementation of these measures also results in substantial impairment in physical and psychological well-being [3]. There is a strong need for an effective vaccine to keep COVID-19 under control.

Development of the COVID-19 vaccines is on the way. According to the World Health Organization, there are 34 and 142 candidate vaccines in clinical and preclinical evaluation, respectively, as of September 3, 2020; four Chinese candidate vaccines have entered phase III clinical trials [4]. The interim analysis of the phase III trials of a China candidate COVID-19 vaccine (Beijing Institute of Biological Product's inactivated vaccine Sinopharm) showed that it had 86% vaccine efficacy against COVID-19 [5]. The analysis also showed the vaccine had a 99% seroconversion rate of neutralizing antibodies and a 100% effectiveness in preventing moderate and severe cases of the disease with no serious safety concern [5]. The United Arab Emirates Ministry of Health and Prevention announced the official registration of the vaccine on December 9, 2020 [5]. On July 22, 2020, the National Health Commission of the People's Republic of China authorized the emergency use of the COVID-19 vaccination and provided COVID-19 vaccines to workers, students, and diplomatic personnel who need to travel aboard, as well as health care workers and personnel working for pandemic and border control [6,7]. According to the recent press release, there were 56,000 Chinese people who had received COVID-19 vaccines developed by Sinopharm before travelling aboard. So far, none of them reported a SARS-CoV-2 infection [8]. It was estimated that at least one COVID-19 vaccine would become available in China by the end of 2020 [9-11]. The market rate will be set at around ¥1000 (about US $154) [10].

The effectiveness of pandemic vaccination campaigns is dependent on both the vaccines' effectiveness and people's willingness to be vaccinated. Simulation experiments have shown that, when the reproduction number (R_0) was 2.5 and vaccination occurred when 5% of the population had been exposed to SARS-CoV-2, a vaccine efficacy of 80% with 75% coverage could reduce the total number of SARS-CoV-2 cases by 85% without any other measures such as social distancing [12]. A systematic review showed that respondents' willingness to receive a H1N1 influenza A vaccine ranged from 8%-67% [13]. Several factors consistently predict the behavioral intention to receive such a vaccine, including risk of infection, severity of the public health event and personal consequences from the illness, harm or adverse events from the vaccination, use of previous vaccination, and ethnicity [13]. To our knowledge, at least 11 published studies and preprints have investigated behavioral intention to receive a COVID-19 vaccination [14-24]. The impact of COVID-19 on the general population and specific groups is quite different, which may cause different responses to COVID-19 vaccination. For example, health care workers' risk of COVID-19 infection was 9-11 times higher than the general population [25], and they had the highest priority to receive a COVID-19 vaccination [26]. Free vaccines are likely to be provided to them in the near future. However, prevalence of poor mental health status (eg, depression, anxiety, or posttraumatic stress disorders) caused by a high risk of infection, being overworked, frustration, discrimination, social isolation, and exhaustion were much higher in this group as compared to the general population [27]. This might explain the lower intention to receive COVID-19 vaccination among health care workers (63.0%-76.4%) [23,24] as compared to that of the general population (57.6%-94.3%) [14-17,20-22]. Factory workers who resumed work during the pandemic are another subpopulation at higher risk of COVID-19 infection than that of the general population, as many factories are crowded settings, making physical distancing challenging [28]. COVID-19 outbreaks in the workplace have been reported in China and other countries [28-30]. Moreover, most of the Chinese factory workers are young. Even if infected with COVID-19, many of them may be asymptomatic and unaware of their infection; they may become a driving force of COVID-19 transmission in the workplace and community [31,32]. Previous studies have reported that the prevalence of poor mental health status was similar for factory workers compared to that of the general population [11] and was much lower than health care workers [27]. It is expected that they have a higher motivation to receive a COVID-19 vaccination than the general population and health care workers. COVID-19 vaccination for factory workers is important to achieve a balance of work resumption and pandemic control.

Health promotion is needed even when free vaccinations are available. To develop effective health promotion campaigns, it is important to understand the facilitators and barriers of COVID-19 vaccination uptake among factory workers. Previous studies have found a number of factors associated with behavioral intention to receive a COVID-19 vaccination for the general population or health care workers, including sociodemographics (eg, age, gender, marital status, income, and history of influenza vaccination), presence of comorbid conditions, and trust in government. Moreover, perceptions related to COVID-19 (eg, risk of infection) and COVID-19 vaccination (eg, perceived efficacy, concerns of side effects, other's acceptance, and confidence to receive vaccination)

influenced their intention to receive a COVID-19 vaccination [14-24]. We considered these factors in this study. Theory-based interventions are more effective than those that are not [33]. In this study, we applied the theory of planned behavior (TPB) as the theoretical framework [34]. The TPB postulates that behavioral intention to adopt a health-related behavior (eg, uptake of a COVID-19 vaccination) is a strong predictor of actual behavior. To form such an intention, one would evaluate the pros and cons of the behavior (positive and negative attitudes), consider whether their significant others would support such behavior (perceived subjective norm), and appraise how much control one has over the behavior (perceived behavioral control) [34]. In recent studies, the TPB has been used to successfully explain behavioral intention and actual behaviors to receive human papillomavirus (HPV) and influenza vaccinations [35-37].

Across countries, it is common to encounter vaccination-related information on social media [38]. Previous studies have shown that over 60% of people in the United States used social media as a common source of information related to HPV and influenza vaccinations [39,40]. During the pandemic, people are also actively seeking information about COVID-19 vaccination on different social media platforms [41]. Several studies have reported that social media use would influence perceptions and behaviors related to vaccination. Four studies found a negative influence of social media on users' perceptions related to vaccination (eg, increase doubt, fear, or barriers for vaccination) [42-45]. Regarding vaccination uptake, women in the United Kingdom who used social media to gather information reported lower pertussis vaccination uptake during pregnancy [46], while positive associations between information exposure through social media and vaccination were found among White and African American adults in the United States [40] and older adults in China [47]. Moreover, different contents related to COVID-19 may have varying effects on personal preventive measures [48]. In this study, we investigated the associations between exposure to different content related to COVID-19 vaccination on social media and behavioral intention to receive a vaccination.

To the best of our knowledge, there have been no studies investigating behavioral intention to receive a COVID-19 vaccination and associated factors among factory workers who resumed work during the COVID-19 pandemic. To address these gaps, this study investigated behavioral intention to receive a self-financed or free COVID-19 vaccination among a sample of factory workers in Shenzhen, China. We examined the effects of factors including sociodemographics, perceptions related to COVID-19 vaccination based on the TPB, exposure to COVID-19 specific information through different media, and COVID-19 preventive measures implemented by individuals and factories.

Methods

Study Design

We conducted a cross-sectional closed online survey of 2053 factory workers in Shenzhen, China on September 1-7, 2020. Of the 13 million residents in Shenzhen in 2018, 65.1% were internal migrants and 34.3% were factory workers [49].

Participants and Data Collection

This study was conducted in the Longhua district of Shenzhen. In Shenzhen, the majority of the factories are located in Longhua. As of 2018, there are over 2000 factories and about one million factory workers in Longhua. Participants were full-time employees of factories in Shenzhen that were 18 years or older. In Shenzhen, factory workers are required to receive a physical examination at designated hospitals or the Centre for Disease Control and Prevention (CDC) annually. All five designated hospitals (three public and two private) and the one district CDC providing physical examination services to factory workers in Longhua were our study sites for recruitment. To avoid selection bias, the fieldworkers approached all adults attending these sites for physical examination during the study period. They briefed prospective participants about the study details, confirmed their eligibility, and invited them to join the study. Participants were guaranteed that participation was voluntary, refusal would have no effect on them, the survey would not collect personal contacts or identification, and data would be kept strictly confidential and only be used for research purposes. Verbal consent was obtained instead of written consent to allow participants to maintain anonymity. We developed an online questionnaire using Questionnaire Star, a commonly used online survey platform in China. Quick Response (QR) codes were generated to access the online questionnaire. Prospective participants were asked to scan the QR code on site to complete the survey. Each mobile device was only allowed to access the online questionnaire once to avoid duplicate responses. The participants were asked not to disseminate the QR codes to access the survey to other people. The survey had 66 items (about 15 items per page for four pages), which took about 15 minutes to complete. The Questionnaire Star performed completeness checks before the questionnaire was submitted. Participants were able to review and change their responses through a "Back" button. An e-coupon of ¥10 (US $1.54) was sent to participants upon completion. In case participants did not have internet access or a smartphone, the research team prepared a tablet computer in each study site for them to complete the online survey. All data was stored in the online server of Questionnaire Star and protected by a password. Only the corresponding author had access to the database. Ethics approval was obtained from the Longhua District CDC (reference: 2020001).

Measures

Design of the Questionnaire

A panel consisting of one CDC staff, two public health researchers, a health psychologist, a senior factory manager, and a factory worker was formed to develop the questionnaire used in this study. The questionnaire was pilot-tested among 10 factory workers to assess clarity and readability. These 10 workers did not participate in the actual survey. Based on the workers' comments, the panel revised and finalized the questionnaire.

Background Characteristics

Participants were asked to report on sociodemographics such as age, gender, relationship status, whether they had a child, highest education level, monthly personal income, status as frontline workers or management, and type of factory they were working in. In addition, participants were also asked about history of seasonal influenza vaccination and whether they had a family member with a history of COVID-19.

Behavioral Intention to Receive COVID-19 Vaccination Under Different Scenarios

Participants were briefed with the following: "COVID-19 vaccines developed by China are likely to become available by the end of 2020." We assessed behavioral intention to receive COVID-19 vaccination under four scenarios: (1) conditional on 50% vaccine efficacy and market rate (¥1000 or US $154), (2) conditional on 80% vaccine efficacy and market rate, (3) conditional on 50% vaccine efficacy and free vaccines, and (4) conditional on 80% vaccine efficacy and free vaccines. On June 2020, the US Food and Drug Administration released guidance for development and licensure of vaccines to prevent COVID-19, which stated that the primary efficacy end point estimate for a placebo-controlled efficacy trial for a COVID-19 vaccine should be at least 50% to ensure that a widely deployed vaccine is effective [50]. Therefore, we chose a threshold of 50% as the lowest estimate of COVID-19 vaccine efficacy in this study. Another study also measured behavioral intention to receive a COVID-19 vaccination conditional on 50% vaccine efficacy [18]. Based on the results of the simulation experiments mentioned in the previous paragraph, Bartsch et al [12] concluded that the vaccine has to have an efficacy of at least 80% to extinguish the COVID-19 epidemic without any other measures. Therefore, we chose a threshold of 80% as an optimal estimation of vaccine efficacy. The threshold of 80% was also close to the vaccine efficacy of the China candidate vaccine (86%) in phase III clinical trials [5]. The cost of COVID-19 vaccines was based on available information in the press release [10].

The response categories were 1 (very unlikely), 2 (unlikely), 3 (neutral), 4 (likely), and 5 (very likely). Behavioral intention was defined as "likely" or "very likely." This definition has been commonly used in previous studies [51-53]. In this study, we measured behavioral intention to receive a COVID-19 vaccination under the condition that its efficacy and cost was made known to the participants. The process ensured that all participants received uniform information and, hence, allowed for better interpretation of the results.

Perceptions Related to COVID-19 Vaccination Based on the TPB

Three scales were constructed to assess perceptions related to COVID-19 vaccination based on the TPB. They were (1) the five-item Positive Attitude Scale (eg, *COVID-19 vaccination is highly effective in protecting you from COVID-19*), (2) the four-item Negative Attitude Scale (eg, *COVID-19 vaccines will have severe side effects*), and (3) the two-item Perceived Subjective Norm Scale (perceived support from doctors/nurses and family members/friends; response categories: 1, disagree; 2, neutral; and 3, agree). The Cronbach α of these scales ranged from .67 to .85; single factors were identified by exploratory factor analysis, explaining for 50.7%-54.0% of total variance. In addition, perceived behavioral control to receive a COVID-19 vaccination was measured by a single item (*receiving a COVID-19 vaccination is easy for you if you want to*; 1, disagree; 2, neutral; and 3, agree).

Influence of Social Media

Participants were asked to report the frequency of their exposure to the following information related to a COVID-19 vaccination on social media (WeChat, WeChat moments, Weibo, Tiktok, etc) in the past month (response categories: 1, almost never; 2, seldom; 3, sometimes; 4, always). Such information included positive information related to COVID-19 vaccination (eg, new vaccines entering clinical trials), negative information related to COVID-19 vaccination (eg, concerns about efficacies and supplies), testimonials given by participants of the COVID-19 clinical trials, and negative information about vaccine incidents in China (eg, selling problematic vaccines and severe side effects).

COVID-19 Preventive Measures Implemented by Individuals and Factories

Participants were asked to report frequency of wearing face masks when having close contact with others in a workplace and other public settings (public spaces or transportation) in the past month (response categories: every time, often, sometimes, never). Participants also reported frequency of sanitizing hands using soaps, liquid soaps, or alcohol-based hand rubs after returning from public spaces or touching public installations or equipment and whether they avoided social or meal gatherings with people who they do not live with and crowded places in the past month. The Shenzhen government advocated that eight preventive measures should be implemented in the factories, including (1) prohibiting nonemployees from entering workplaces, (2) taking body temperature and sanitizing hands for all employees before entering the workplace, (3) providing face masks to all employees, (4) keeping adequate distance (eg, >1 meter) between workstations, (5) requiring employees to wear face masks when they have close contact with other people, (6) disinfecting the workplace frequently, (7) maintaining adequate ventilation in the workplace, and (8) setting up partitions in factory canteens [54,55]. Participants reported whether their factory implemented these eight preventive measures. A composite indicator variable was constructed by counting the number of preventive measures implemented by the factory (ranging from 0 to 8). The English and Chinese versions of the questionnaire are shown in Multimedia Appendix 1.

Sample Size Planning

The target sample size was 2000. Given a statistical power of 0.80 and an alpha value of .05, and assuming the level of behavioral intention to receive a COVID-19 vaccination in the reference group (without a facilitating condition) to be 30%-70%, the sample size could detect a smallest odds ratio (OR) of 1.29 between those with and without such facilitating condition (PASS 11.0; NCSS, LLC).

Statistical Analysis

The binary variables on behavioral intention to receive a COVID-19 vaccination conditional on 80% vaccine efficacy and market rate and conditional on same efficacy and free vaccines were used as the dependent variables. Multilevel logistic regression models (level 1: study sites; level 2: individual participants) were used to analyze factors associated with the dependent variables. Random intercept models were used to allow the intercept of the regression model to vary across study sites, which could account for intracorrelated nested data. Multilevel logistic regression models are commonly used in studies with similar sampling methods [28,56]. A univariate two-level logistic regression model first assessed the significance of the association between each of the background characteristics and the dependent variables. Background characteristics with $P<.05$ in univariate analysis were adjusted in the multivariate two-level logistic regression model.

In addition, using behavioral intention to receive a COVID-19 vaccination conditional on 80% vaccine efficacy and market rate and conditional on same efficacy and free vaccines were used as ordinal dependent variables (from 1 to 5), and background characteristics were used as independent variables; ORs were obtained using ordinal logistic regressions. Adjustment for significant background characteristics, associations between independent variables of interest (perceptions, information exposure through social media, and preventive measures implemented by individuals and factories) and the dependent variables were then assessed by adjusted odds ratios (AORs). A similar approach was used in previous studies [57]. Principal component analysis with varimax rotation was used to perform explanatory factor analysis. Correlations between information exposure through social media and perceptions related to COVID-19 vaccination were also investigated. Pearson correlation coefficients (r) were obtained. SPSS version 26.0 (IBM Corp) was used for data analysis, with $P<.05$ considered statistically significant.

Results

Background Characteristics

Out of 2653 eligible factory workers (between 60 and 1200 across study sites) that were approached, 2053 completed the online survey (between 40 and 968 across study sites). The overall response rate was 77.4% (ranging from 66.7% to 80.7% at different sites). Main reasons for nonresponse were lack of time and other logistic reasons. All participants that were approached had access to the internet or a smartphone, and none of them used the tablet computers prepared by the research team. Over half of the participants were younger than 40 years (n=1490, 72.6%), were female (n=1179, 57.4%), were married (n=1455, 70.9%), had children (n=1466, 71.4%), did not receive tertiary education (n=1472, 71.7%), had a monthly income less than ¥5000 (US $773; n=1421, 69.2%), were frontline workers (n=1476, 71.9%), and were working for electronic device manufacturers (n=1473, 71.7%). Among the participants, 20.3% (n=416) had received a seasonal influenza vaccination at least once, and 1.6% (n=32) had at least one family member with a history of COVID-19 (Table 1).

Table 1. Background characteristics of the participants (N=2053).

Characteristics	Participants, n (%)
Age group (years)	
18-30	757 (36.9)
31-40	733 (35.7)
41-50	491 (23.9)
>50	72 (3.5)
Gender	
Male	874 (42.6)
Female	1179 (57.4)
Relationships status	
Currently single	448 (21.8)
Having a stable boyfriend/girlfriend	150 (7.3)
Married	1455 (70.9)
Have children	
No	587 (28.6)
Yes	1466 (71.4)
Highest education level attained	
Junior high or below	859 (41.8)
Senior high or equivalent	613 (29.9)
College/university or above	581 (28.3)
Monthly personal income (¥; US $)	
<3000 (463.84)	512 (24.9)
3000-4999 (463.84-772.92)	909 (44.3)
5000-6999 (773.07-1082.15)	359 (17.5)
7000-9999 (1082.30-1545.99)	150 (7.3)
≥10,000 (1546.14)	123 (6.0)
Type of work	
Frontline workers	1476 (71.9)
Management staff	577 (28.1)
Factory type	
Electronic devices manufacturers	1473 (71.7)
Other factories	580 (28.3)
History of seasonal influenza vaccination	
No	1637 (79.7)
Yes	416 (20.3)
Having at least one family member with a history of COVID-19	
No	2021 (98.4)
Yes	32 (1.6)

Behavioral Intention to Receive COVID-19 Vaccination Under Different Scenarios

The prevalence of behavioral intention to receive a COVID-19 vaccination was 53.5% (n=1099, conditional on 50% vaccine efficacy and market rate), 66.6% (n=1368, conditional on 80% vaccine efficacy and market rate), 75.6% (n=1551, conditional on 50% vaccine efficacy and free vaccines), and 80.6% (n=1655, conditional on 80% vaccine efficacy and free vaccines; Table 2).

Table 2. Perceptions related to COVID-19 vaccination and preventive measures taken up by participants and the factories they were working in (N=2053).

Perceptions	Participants, n (%)	Mean (SD)
Behavioral intention to take up COVID-19 vaccination		
Intention to get COVID-19 vaccines conditional on 50% efficacy and market rate (¥1000 or US $140)		N/A[a]
Very unlikely	222 (10.8)	
Unlikely	208 (10.1)	
Neutral	524 (25.5)	
Likely	543 (26.4)	
Very likely	556 (27.1)	
Intention to get COVID-19 vaccines conditional on 80% efficacy and market rate (¥1000 or US $140)		N/A
Very unlikely	155 (7.5)	
Unlikely	147 (7.2)	
Neutral	383 (18.7)	
Likely	639 (31.1)	
Very likely	729 (35.5)	
Intention to get COVID-19 vaccines conditional on 50% efficacy and free vaccines		N/A
Very unlikely	108 (5.3)	
Unlikely	98 (4.8)	
Neutral	296 (14.4)	
Likely	453 (22.1)	
Very likely	1098 (53.5)	
Intention to get COVID-19 vaccines conditional on 80% efficacy and free vaccines		N/A
Very unlikely	117 (5.7)	
Unlikely	77 (3.8)	
Neutral	204 (9.9)	
Likely	320 (15.6)	
Very likely	1335 (65.0)	
Perceptions related to COVID-19 vaccination based on the theory of planned behavior		
Positive attitudes toward COVID-19 vaccination	N/A	
Positive Attitude Scale[b]		13.3 (2.3)
Negative attitudes toward COVID-19 vaccination	N/A	
Negative Attitude Scale[c]		8.0 (1.9)
Perceived subjective norm related to COVID-19 vaccination	N/A	
Perceived Subjective Norm Scale[d]		5.1 (1.2)
Perceived behavioral control to receive COVID-19 vaccination	N/A	
Receiving COVID-19 vaccination is easy for you if you want to		2.3 (0.7)
Influence of social media related to COVID-19 vaccination		
Frequency of exposure to positive information related to COVID-19 vaccination (eg, new vaccines entering clinical trials, promising efficacies of the vaccines, and vaccines will enter the market soon) on social media	N/A	2.8 (1.0)
Frequency of exposure to negative information related to COVID-19 vaccination (eg, concerns about efficacies and supplies, side effects of the vaccines, and receiving vaccines will cause COVID-19) on social media	N/A	2.3 (0.9)
Frequency of exposure to testimonials given by participants of the COVID-19 vaccine clinical trials on social media	N/A	1.9 (1.0)

Perceptions	Participants, n (%)	Mean (SD)
Frequency of exposure to negative information about other vaccines in China (eg, selling problematic vaccines and severe side effects) on social media	N/A	2.0 (1.0)
Personal COVID-19 preventive measures in the past month		
Frequency of wearing a face mask in public places/transportation other than the workplace		N/A
Every time	1675 (81.6)	
Often	280 (13.6)	
Sometimes	82 (4.0)	
Never	16 (0.8)	
Frequency of wearing a face mask when you have close contact with other people in the workplace		N/A
Every time	1519 (74.0)	
Often	370 (18.0)	
Sometimes	144 (7.0)	
Never	20 (1.0)	
Self-reported sanitizing hands (using soaps, liquid soaps, or alcohol-based sanitizer) after returning from public spaces or touching public installation		N/A
Every time	1217 (59.3)	
Often	495 (24.1)	
Sometimes	323 (15.7)	
Never	18 (0.9)	
Self-reported avoiding social gatherings with other people who do not live together		N/A
No	1165 (56.7)	
Yes	888 (43.3)	
Self-reported avoiding crowded places		N/A
No	1309 (63.8)	
Yes	744 (36.2)	
Number of preventive measures implemented by the factory	N/A	6.1 (2.5)

[a]N/A: not applicable.

[b]Positive Attitude Scale, 5 items, Cronbach alpha: .84; 1 factor was identified by exploratory factor analysis, explaining for 54.0% of total variance.

[c]Negative Attitude Scale, 4 items, Cronbach alpha: .67; 1 factor was identified by exploratory factor analysis, explaining for 50.9% of total variance.

[d]Perceived Subjective Norm Scale, 2 items, Cronbach alpha: .85.

Perceptions Related to COVID-19 Based on the TPB and Influence of Social Media

Means and SDs of items and scales related to COVID-19 vaccination based on the TPB are described in Table 2 and Multimedia Appendix 2. Among the participants, 66.4% (n=1363) were sometimes or always exposed to positive information related to COVID-19 vaccinations in the past month, while fewer participants were sometimes or always exposed to negative information related to COVID-19 vaccinations (n=842, 41.0%) or other vaccines in China (n=571, 27.8%), or exposed to testimonials given by participants of COVID-19 vaccination clinical trials (n=594, 28.9%).

COVID-19 Preventive Measures Implemented by Individuals and Factories

In the past month, 74.0% (n=1519) and 81.6% (n=1675) of participants reported wearing a face mask every time they had close contact with other people in the workplace and in other public settings, respectively. More than half of the participants self-reported sanitizing hands (n=1217, 59.3%), avoiding social or meal gatherings (n=1165, 56.7%), and avoiding crowded places (n=1309, 63.8%; Table 2 and Multimedia Appendix 2).

Factors Associated With Behavioral Intention to Receive a COVID-19 Vaccination

In the univariate logistic regression analysis, age group, relationship status, having children, education level, monthly personal income, status as frontline workers or management staff, history of seasonal influenza vaccination, and having a family member with a history of COVID-19 were significantly associated with one or both dependent variables (Table 3).

Table 3. Associations between background characteristics and behavioral intention to receive COVID-19 vaccination under different scenarios (N=2053).

Characteristics	Conditional on 80% efficacy and market rate, OR[a,b] (95% CI)	Conditional on 80% efficacy and free vaccines, OR (95% CI)
Age group (years)		
18-30	1.0	1.0
31-40	0.80 (0.64-1.01)	0.95 (0.72-1.25)
41-50	0.62 (0.48-0.79)***	0.61 (0.42-0.76)***
>50	0.48 (0.29-0.79)**	0.57 (0.32-1.02)
Gender		
Male	1.0	1.0
Female	1.16 (0.96-1.41)	0.99 (0.78-1.24)
Relationships status		
Currently single	1.0	1.0
Having a stable boyfriend/girlfriend	1.44 (0.94-2.20)	1.55 (0.89-2.71)
Married	0.94 (0.74-1.18)	0.77 (0.58-1.02)
Having children		
No	1.0	1.0
Yes	0.77 (0.62-0.95)*	0.66 (0.50-0.86)*
Highest education level attained		
Junior high or below	1.0	1.0
Senior high or equivalent	1.57 (1.26-1.96)***	2.09 (1.60-2.73)***
College/university or above	1.94 (1.52-2.47)***	3.39 (2.69-5.01)***
Monthly personal income (¥; US $)		
<3000 (463.84)	1.0	1.0
3000-4999 (463.84-772.92)	1.11 (0.89-1.40)	1.20 (0.89-1.50)
5000-6999 (773.07-1082.15)	1.36 (1.01-1.83)*	1.39 (0.99-1.97)*
7000-9999 (1082.30-1545.99)	1.47 (0.97-2.21)*	1.76 (1.05-2.94)*
≥10,000 (1546.14)	1.60 (1.01-2.56)*	4.82 (2.05-11.36)***
Type of work		
Frontline workers	1.0	1.0
Management staff	1.24 (1.01-1.54)**	1.66 (1.26-2.19)***
Factory type		
Electronic devices manufacturer	1.0	1.0
Other factories	1.27 (0.97-1.52)	1.09 (0.83-1.42)
History of seasonal influenza vaccination		
No	1.0	1.0
Yes	1.29 (1.02-1.64)*	1.28 (0.95-1.71)
Having a family member with history of COVID-19		
No	1.0	1.0
Yes	4.85 (2.21-10.53)***	6.49 (3.10-13.51)***

[a]OR: odds ratio.

[b]Crude ORs obtained by using univariate two-level logistic regression models.

*$P<.05$, **$P<.01$, ***$P<.001$.

After adjusting for these significant background characteristics, positive attitudes toward COVID-19 vaccination (AOR 1.20, 95% CI 1.15-1.25 and AOR 1.24, 95% CI 1.19-1.30), perceived support from significant others on COVID-19 vaccination uptake (AOR 1.43, 95% CI 1.32-1.55 and AOR 1.37, 95% CI 1.25-1.50), and perceived behavioral control to get a COVID-19 vaccination (AOR 1.51, 95% CI 1.32-1.73 and AOR 1.28, 95% CI 1.09-1.51) were positively associated with both dependent variables (dependent on 80% efficacy and market rate vaccines and dependent on 80% efficacy and free vaccines, respectively). Regarding social media influence, higher frequency of exposure to positive information related to a COVID-19 vaccination was associated with higher intention to receive a COVID-19 vaccination at market rate (AOR 1.53, 95% CI 1.39-1.70) or to receive a free vaccination (AOR 1.52, 95% CI 1.35-1.71). Higher self-reported compliance with wearing a face mask in the workplace (AOR 1.27, 95% CI 1.02-1.58 and AOR 1.67, 95% CI 1.24-2.27) and other public spaces (AOR 1.80, 95% CI 1.42-2.29 and AOR 1.34, 95% CI 1.01-1.77), hand hygiene (AOR 1.21, 95% CI 1.00-1.47 and AOR 1.52, 95% CI 1.19-1.93), and avoiding social and meal gatherings (AOR 1.22, 95% CI 1.01-1.47 and AOR 1.55, 95% CI 1.23-1.95) and crowded places (AOR 1.24, 95% CI 1.02-1.51 and AOR 1.73, 95% CI 1.37-2.18) were also positively associated with one or both dependent variables (dependent on 80% efficacy and market rate vaccines and dependent on 80% efficacy and free vaccines, respectively). A higher number of COVID-19 preventive measures implemented by the factory were significantly associated with a higher intention to receive COVID-19 vaccination under both scenarios (AOR 1.08, 95% CI 1.04-1.12 and AOR 1.06, 95% CI 1.01-1.11, respectively; Table 4).

Table 4. Factors associated with behavioral intention to receive a COVID-19 vaccination under different scenarios (N=2053).

Factors	Conditional on 80% and market rate, AOR[a,b] (95% CI)	Conditional on 80% efficacy and free vaccines, AOR (95% CI)
Perceptions relate to COVID-19 vaccination based on the theory of planned behavior		
Positive Attitude Scale	1.20 (1.15-1.25)***	1.24 (1.19-1.30)***
Negative Attitude Scale	0.98 (0.93-1.03)	1.00 (0.94-1.06)
Perceived Subjective Norm Scale	1.43 (1.32-1.55)***	1.37 (1.25-1.50)***
Perceived behavioral control to receive COVID-19 vaccination	1.51 (1.32-1.73)***	1.28 (1.09-1.51)**
Influence of social media related to COVID-19 vaccination		
Frequency of exposure to positive information related to COVID-19 vaccination on social media	1.53 (1.39-1.70)**	1.52 (1.35-1.71)***
Frequency of exposure to negative information related to COVID-19 vaccination on social media	1.11 (0.99-1.23)	1.07 (0.95-1.21)
Frequency of exposure to testimonials given by participants of the COVID-19 vaccine clinical trials on social media	1.10 (0.99-1.21)	1.00 (0.89-1.11)
Frequency of exposure to negative information about vaccine incidents in China on social media	0.95 (0.86-1.04)	0.93 (0.83-1.05)
Personal COVID-19 preventive measures in the past month		
Consistent use of face mask in public places/transportation other than the workplace		
No	1.0	1.0
Yes	1.80 (1.42-2.29)***	1.34 (1.01-1.77)*
Consistent use of face mask when you have close contact with other people in the workplace		
No	1.0	1.0
Yes	1.27 (1.02-1.58)*	1.67 (1.24-2.27)**
Self-reported sanitizing hands (using soaps, liquid soaps, or alcohol-based sanitizer) every time after returning from public spaces or touching public installation		
No	1.0	1.0
Yes	1.21 (1.00-1.47)*	1.52 (1.19-1.93)**
Self-reported avoiding social gatherings with other people who do not live together		
No	1.0	1.0
Yes	1.22 (1.01-1.47)*	1.55 (1.23-1.95)***
Self-reported avoiding crowded places		
No	1.0	1.0
Yes	1.24 (1.02-1.51)*	1.73 (1.37-2.18)***
Number of preventive measures implemented by the factory	1.08 (1.04-1.12)***	1.06 (1.01-1.11)*

[a]AOR: adjusted odds ratios

[b]Background characteristics with $P<.05$ in univariate two-level logistic regression analysis were adjusted in the multivariate two-level logistic regression models.

*$P<.05$, **$P<.01$, ***$P<.001$.

When behavioral intention was treated as ordinal variables (from 1 to 5) and used as dependent variables, the same sets of associated factors were identified by univariate and multivariate ordinal logistic regression models. The results are presented in Multimedia Appendix 3.

Correlation Between Information Exposure Through Social Media and Perceptions Related to COVID-19 Vaccination Based on the TPB

Frequency of exposure to positive information related to COVID-19 vaccinations on social media was positively correlated with positive attitudes ($r=0.083$; $P<.001$), perceived

subjective norm (r=0.101; P<.001), and perceived behavioral control (r=0.064; P=.004) related to a COVID-19 vaccination. A negative correlation was found between social media exposure and negative attitudes toward a COVID-19 vaccination (r=–0.090; P<.001). Moreover, frequency of exposure to positive information related to COVID-19 vaccinations on social media was negatively correlated with positive attitudes (r=–0.080; P<.001), perceived subjective norm (r=–0.107; P<.001), and perceived behavioral control (r=–0.069; P=.002) related to a COVID-19 vaccination. Furthermore, frequency of exposure to testimonials given by COVID-19 vaccine clinical trial participants on social media was negatively correlated with positive attitudes (r=–0.052, P=.02) but was positively correlated with perceived behavioral control to get a COVID-19 vaccination (r=0.062, P=.005). In addition, frequency of exposure to negative information about other vaccines in China on social media was negatively correlated with positive attitudes (r=–0.106, P<.001) and perceived subjective norm (r=–0.132, P<.001) related to a COVID-19 vaccination (Multimedia Appendix 4).

Discussion

Our findings represent one of the latest estimates of COVID-19 vaccination acceptability in China and can be used to project future vaccine uptake. Factory workers' behavioral intention to receive a COVID-19 vaccination was more sensitive to its cost than vaccine efficacy. Given the same vaccine efficacy, the behavioral intention varied from 53.6%-66.6% at market rate to about 80% for a free vaccination. However, small increases were observed (13.1% under market rate and 2% for free vaccination) when comparing behavioral intention conditional on 50% vaccine efficacy and those conditional on 80% vaccine efficacy. The prevalence of behavioral intention conditional on free vaccination was higher than that reported in the United States [16] and Saudi Arabia [22] but was lower than that of the general population in Malaysia [14] and China [23]. A meta-analysis showed that only 43%-62% of those with a behavioral intention would translate that into related actions [58]. Therefore, effective health promotion is needed when COVID-19 vaccines become available to achieve high coverage at a population level.

Our findings provide empirical insights to inform health promotion development. Similar strategies can be used to promote COVID-19 vaccinations at market rate or free vaccinations, as the associated factors of these two dependent variables were similar. Older participants had a lower intention to receive a COVID-19 vaccination. This finding is consistent with a previous study targeting the general population in the United States [20]. Younger people may be more receptive to innovations [59]. Participants with lower education had a lower intention to receive a COVID-19 vaccination compared to workers with higher levels of education. The former might find it difficult to understand information related to COVID-19 vaccination due to a lower literacy level. The finding for income coincides with previous studies that suggested that intention to receive COVID-19 vaccines and uptake of other vaccines were lower among individuals with lower income [20,60]. As compared to frontline workers, management staff were more willing to receive a COVID-19 vaccination. Studies among factory workers have shown that management staff were more likely to adopt personal measures to prevent COVID-19 [28]. Having children was associated with a lower intention to receive a COVID-19 vaccination; future studies should explore whether there are any specific barriers for parents. A history of seasonal influenza vaccination was associated with a higher intention to receive a COVID-19 vaccination at market rate but not with the intention to receive a free vaccination. In China, free seasonal influenza vaccination is provided to older adults, while out-of-pocket payment is required for other groups. Factory workers who had taken a seasonal influenza vaccination may have positive beliefs toward self-financed vaccination. Furthermore, having a family member with history of COVID-19 was strongly associated with a higher intention to receive a COVID-19 vaccination. This finding was expected, as these participants had direct experience related to COVID-19 and were more likely to perceive it as a serious health threat.

The TPB is a potentially useful framework to guide the development of future programs, as three of the four TPB constructs used in this study were significantly associated with both dependent variables in expected directions. It would be useful to increase positive attitudes toward COVID-19 vaccination, as this was found to be a facilitator. In addition to the beneficial effect for oneself (eg, preventing COVID-19 and returning to normal lives), health communication messages should also emphasize that COVID-19 vaccination uptake would result in herd immunization, which could contribute to COVID-19 control. Building up confidence related to the vaccine supply may also be a useful strategy. Over 60% of participants perceived that medical professionals, family, and friends would support them in taking the COVID-19 vaccination. This perception was also a facilitator. Future programs should consider involving the significant others of factory workers to create a subjective norm favoring COVID-19 vaccination uptake. It would also be useful to enhance perceived behavioral control, as this was another facilitator. There is more room for improvement. Outreach in the factories and providing vaccination on-site may be a useful strategy to improve perceived behavioral control among the workers. Relatively few participants had concerns related to cost, side effects, and duration of vaccine protection. The associations between these concerns and behavioral intention were not statistically significant. Addressing these concerns might not be a useful strategy in future promotion campaigns.

Our findings suggest that COVID-19 vaccination triggered intensive responses on social media, as about 60% of the participants sometimes or always were exposed to information specific to COVID-19 vaccination on different social media platforms. Our results showed that exposure to positive information related to COVID-19 vaccination through social media was positively correlated with positive perceptions (ie, positive attitudes, perceived subjective norm, and perceived behavioral control) related to COVID-19 vaccination. These positive perceptions were determinants of behavioral intention to get a COVID-19 vaccination in this study. It is possible that higher amounts of positive information exposure on social media would enhance these positive perceptions, which in turn

increases behavioral intention to get a COVID-19 vaccination. Longitudinal studies are needed to test whether this pathway exists. Negative information about vaccines is uniquely attractive to social media. Studies have shown that a major vaccine incident (Changchun Changsheng) had significantly impaired the confidence of vaccines among Chinese people [61]. However in this study, negative information about these vaccine incidents did not influence participants' behavioral intentions to receive a COVID-19 vaccination.

Factory workers who reported higher compliance to personal preventive measures were more willing to receive a COVID-19 vaccination. These people may have stronger motivation and self-efficacy to protect themselves, and a COVID-19 vaccination is likely to be considered as a useful means for protection. Preventive measures implemented by the factories also played important roles. More measures implemented by the factories was associated with a higher intention to receive a COVID-19 vaccination. Through implementation of these measures, factories could cultivate widely shared organizational norms to facilitate behavioral changes among the workers [28,62].

This study has some limitations. First, a direct measure of perceived behavioral control should assess self-efficacy and perceived controllability [63]. Previous studies have suggested these two constructs were differentially associated with behavioral intention and actual behaviors [63]. Due to the limited length of the questionnaire, we only used a single item to measure perceived behavioral control, which mainly covered self-efficacy. Failure to measure perceived controllability together with self-efficacy was one major limitation of this study. This limitation made this study less comparable to other studies using the TPB. Second, this study focused on factory workers and did not study the general population in Shenzhen. In addition, we only included factory workers in one Chinese city. Generalizations should be made cautiously to individuals working in other places in China. Third, since the study was anonymous and did not collect participants' identification, we were not able to collect the information of those who refused to join the study. Factory workers who refused to join the study might have different characteristics as compared to study participants. Since most of the factory workers in Shenzhen are internal migrants, these is no accurate census data for this group. Therefore, we were not able to perform weighting for our sample; selection bias existed. However, our response rate was relatively high as compared to other online surveys of similar topics [23]. Fourth, data was self-reported and verification was not feasible. Recall bias might exist. Participants may have also overreported their intention and compliance with personal preventive measures due to social desirability. Fifth, most items and scales used in this study were self-constructed based on those from previous studies on H1N1 and seasonal influenza vaccination in China [64,65]. The internal reliability of these scales were acceptable, but these scales may require external validation. Moreover, casual relationships cannot be determined due to the cross-sectional design of this study.

In summary, factory workers in China reported a high behavioral intention to receive a COVID-19 vaccination. The behavioral intention was cost-sensitive, and the proposed market rate was accepted by the majority of the participants. The TPB is a useful framework to guide the development of future campaigns promoting COVID-19 vaccination in this group.

Acknowledgments

This study was funded by the Key Discipline of Infectious Diseases Control and Prevention of Long Hua (grant number 2020-2014).

Conflicts of Interest

None declared.

Multimedia Appendix 1
The English and Chinese versions of the questionnaire used in the online survey.
[DOCX File , 26 KB - jmir_v23i3e24673_app1.docx]

Multimedia Appendix 2
Item responses of perceptions related to COVID-19 vaccination and preventive measures taken up by participants and the factories they were working in.
[DOCX File , 17 KB - jmir_v23i3e24673_app2.docx]

Multimedia Appendix 3
Factors associated with behavioral intention to receive a COVID-19 vaccination obtained by using univariate and multivariate ordinal logistic regression models.
[DOCX File , 19 KB - jmir_v23i3e24673_app3.docx]

Multimedia Appendix 4
Correlations between information exposure through social media and perceptions related to COVID-19 vaccination based on the theory of planned behavior.
[DOCX File , 16 KB - jmir_v23i3e24673_app4.docx]

References

1. Coronavirus disease (COVID-19) weekly epidemiological update and weekly operational update. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-11-17]
2. COVID-19 sends most G20 members into negative GDP growth, except China. The State Council the People's Republic of China. 2020. URL: http://english.www.gov.cn/news/topnews/202009/05/content_WS5f5398c8c6d0f7257693b957.html [accessed 2020-09-23]
3. Carvalho Aguiar Melo M, de Sousa Soares D. Impact of social distancing on mental health during the COVID-19 pandemic: an urgent discussion. Int J Soc Psychiatry 2020 Sep;66(6):625-626 [FREE Full text] [doi: 10.1177/0020764020927047] [Medline: 32438880]
4. Draft landscape and tracker of COVID-19 candidate vaccines. World Health Organization. 2020. URL: http://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines [accessed 2020-09-23]
5. UAE Ministry of Health and Prevention announces official registration of inactivated COVID-19 vaccine used in #4Humanity Trials. Emirates News Agency. 2020. URL: http://www.wam.ae/en/details/1395302893589 [accessed 2020-12-10]
6. The Vaccine Administration Law of the People's Republic of China. Ministry of National Defense of the People's Republic of China. 2019. URL: http://www.mod.gov.cn/big5/regulatory/2019-06/29/content_4844849.htm [accessed 2020-12-02]
7. Press release on October 20, 2020. National Health Commission of the People's Republic of China. 2020. URL: http://www.nhc.gov.cn/xcs/fkdt/202010/a95d956c39cf4393b400b42aa8433033.shtml [accessed 2020-12-02]
8. Sinopharm: 56000 people received COVID-19 vaccines before travelling aboard, none of them got infected. Sina News. 2020. URL: https://news.sina.com.cn/c/2020-11-07/doc-iiznctke0022556.shtml [accessed 2020-12-08]
9. Are vaccines ready for the influenza plus COVID-19? CCTV News. 2020. URL: https://m.news.cctv.com/2020/09/15/ARTIF5GfsYVC53DeGRZZtSNF200915.shtml [accessed 2020-09-23]
10. Brant R. Coronavirus: vaccine front-runner China already inoculating workers. BBC News. 2020. URL: http://www.bbc.com/news/world-asia-china-53917315 [accessed 2020-09-23]
11. Chinese COVID-19 vaccines expected by year end: expert. Global Times. 2020. URL: http://www.globaltimes.cn/content/1197848.shtml [accessed 2020-09-23]
12. Bartsch SM, O'Shea KJ, Ferguson MC, Bottazzi ME, Wedlock PT, Strych U, et al. Vaccine efficacy needed for a COVID-19 coronavirus vaccine to prevent or stop an epidemic as the sole intervention. Am J Prev Med 2020 Oct;59(4):493-503 [FREE Full text] [doi: 10.1016/j.amepre.2020.06.011] [Medline: 32778354]
13. Nguyen T, Henningsen KH, Brehaut JC, Hoe E, Wilson K. Acceptance of a pandemic influenza vaccine: a systematic review of surveys of the general public. Infect Drug Resist 2011;4:197-207. [doi: 10.2147/IDR.S23174] [Medline: 22114512]
14. Wong LP, Alias H, Wong P, Lee HY, AbuBakar S. The use of the health belief model to assess predictors of intent to receive the COVID-19 vaccine and willingness to pay. Hum Vaccin Immunother 2020 Sep 01;16(9):2204-2214. [doi: 10.1080/21645515.2020.1790279] [Medline: 32730103]
15. Wang J, Jing R, Lai X, Zhang H, Lyu Y, Knoll MD, et al. Acceptance of COVID-19 vaccination during the COVID-19 pandemic in China. Vaccines (Basel) 2020 Aug 27;8(3):482 [FREE Full text] [doi: 10.3390/vaccines8030482] [Medline: 32867224]
16. Reiter PL, Pennell ML, Katz ML. Acceptability of a COVID-19 vaccine among adults in the United States: how many people would get vaccinated? Vaccine 2020 Sep 29;38(42):6500-6507 [FREE Full text] [doi: 10.1016/j.vaccine.2020.08.043] [Medline: 32863069]
17. Thunstrom L, Ashworth M, Finnoff D, Newbold S. Hesitancy towards a COVID-19 vaccine and prospects for herd immunity. SSRN J. Preprint posted online May 11, 2020. [FREE Full text] [doi: 10.2139/ssrn.3593098]
18. Harapan H, Wagner AL, Yufika A, Winardi W, Anwar S, Gan AK, et al. Acceptance of a COVID-19 vaccine in Southeast Asia: a cross-sectional study in Indonesia. Front Public Health 2020;8:381. [doi: 10.3389/fpubh.2020.00381] [Medline: 32760691]
19. García LY, Cerda AA. Contingent assessment of the COVID-19 vaccine. Vaccine 2020 Jul 22;38(34):5424-5429 [FREE Full text] [doi: 10.1016/j.vaccine.2020.06.068] [Medline: 32620375]
20. Fisher KA, Bloomstone SJ, Walder J, Crawford S, Fouayzi H, Mazor KM. Attitudes toward a potential SARS-CoV-2 vaccine: a survey of U.S. adults. Ann Intern Med 2020 Dec 15;173(12):964-973 [FREE Full text] [doi: 10.7326/M20-3569] [Medline: 32886525]
21. COCONEL Group. A future vaccination campaign against COVID-19 at risk of vaccine hesitancy and politicisation. Lancet Infect Dis 2020 Jul;20(7):769-770 [FREE Full text] [doi: 10.1016/S1473-3099(20)30426-6] [Medline: 32445713]
22. Padhi BK, Al-Mohaithef M. Determinants of COVID-19 vaccine acceptance in Saudi Arabia: a web-based national survey. medRxiv. Preprint posted online May 30, 2020. [FREE Full text] [doi: 10.1101/2020.05.27.20114413]
23. Fu C, Wei Z, Pei S, Li S, Sun X, Liu P. Acceptance and preference for COVID-19 vaccination in health-care workers (HCWs). medRxiv. Preprint posted online April 14, 2020. [FREE Full text] [doi: 10.1101/2020.04.09.20060103]
24. Kwok KO, Li KK, Wei WI, Tang A, Wong SYS, Lee SS. Are we ready when COVID-19 vaccine is available? Study on nurses' vaccine hesitancy in Hong Kong. medRxiv. Preprint posted online July 17, 2020. [FREE Full text] [doi: 10.1101/2020.07.17.20156026]

25. COVID-19 Scientific Advisory Group rapid response report. Alberta Health Services. 2020. URL: http://www.albertahealthservices.ca/assets/info/ppih/if-ppih-covid-19-hcw-risk-rapid-review.pdf [accessed 2020-12-11]
26. How CDC is making COVID-19 vaccine recommendations. Centers for Disease Control and Prevention. 2020. URL: http://www.cdc.gov/coronavirus/2019-ncov/vaccines/recommendations-process.html [accessed 2020-12-11]
27. Preti E, Di Mattei V, Perego G, Ferrari F, Mazzetti M, Taranto P, et al. The psychological impact of epidemic and pandemic outbreaks on healthcare workers: rapid review of the evidence. Curr Psychiatry Rep 2020 Jul 10;22(8):43 [FREE Full text] [doi: 10.1007/s11920-020-01166-z] [Medline: 32651717]
28. Pan Y, Fang Y, Xin M, Dong W, Zhou L, Hou Q, et al. Self-reported compliance with personal preventive measures among Chinese factory workers at the beginning of work resumption following the COVID-19 outbreak: cross-sectional survey study. J Med Internet Res 2020 Sep 29;22(9):e22457 [FREE Full text] [doi: 10.2196/22457] [Medline: 32924947]
29. Coronavirus: Banham Poultry factory Covid 19 outbreak 'contained'. BBC News. 2020. URL: http://www.bbc.com/news/uk-england-norfolk-54056299 [accessed 2020-09-29]
30. Vickers T, Shipton H, Wong W. Greencore factory: timeline of a coronavirus outbreak shows staff must be listened to. The Conversation. 2020. URL: https://theconversation.com/greencore-factory-timeline-of-a-coronavirus-outbreak-shows-staff-must-be-listened-to-145590 [accessed 2020-09-29]
31. Yonker L, Neilan A, Bartsch Y, Patel A, Regan J, Arya P, et al. Pediatric severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): clinical presentation, infectivity, and immune responses. J Pediatr 2020 Dec;227:45-52.e5 [FREE Full text] [doi: 10.1016/j.jpeds.2020.08.037] [Medline: 32827525]
32. Virtual press conference on COVID-19 in the Western Pacific. World Health Organization. 2020. URL: http://www.who.int/westernpacific/news/speeches/detail/virtual-press-conference-on-covid-19-in-the-western-pacific [accessed 2020-09-29]
33. Michie S, Johnston M, Francis J, Hardeman W, Eccle M. From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. App Psychol 2008;57(4):660-680. [doi: 10.1111/j.1464-0597.2008.00341.x]
34. Ajzen I. The theory of planned behavior. Organizational Behav Hum Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
35. Catalano HP, Knowlden AP, Birch DA, Leeper JD, Paschal AM, Usdan SL. Using the theory of planned behavior to predict HPV vaccination intentions of college men. J Am Coll Health 2017 Apr;65(3):197-207. [doi: 10.1080/07448481.2016.1269771] [Medline: 27960609]
36. Wang LD, Lam WWT, Fielding R. Determinants of human papillomavirus vaccination uptake among adolescent girls: a theory-based longitudinal study among Hong Kong Chinese parents. Prev Med 2017 Sep;102:24-30. [doi: 10.1016/j.ypmed.2017.06.021] [Medline: 28652087]
37. Wu AM, Lau JT, Ma Y, Cheng K, Lau MM. A longitudinal study using parental cognitions based on the theory of planned behavior to predict childhood influenza vaccination. J Infect Public Health 2020 Jul;13(7):970-979 [FREE Full text] [doi: 10.1016/j.jiph.2020.04.009] [Medline: 32418882]
38. Burki T. Vaccine misinformation and social media. Lancet Digital Health 2019 Oct;1(6):e258-e259. [doi: 10.1016/s2589-7500(19)30136-0]
39. Dilley SE, Peral S, Straughn JM, Scarinci IC. The challenge of HPV vaccination uptake and opportunities for solutions: lessons learned from Alabama. Prev Med 2018 Aug;113:124-131. [doi: 10.1016/j.ypmed.2018.05.021] [Medline: 29800594]
40. Ahmed N, Quinn SC, Hancock GR, Freimuth VS, Jamison A. Social media use and influenza vaccine uptake among White and African American adults. Vaccine 2018 Nov 26;36(49):7556-7561. [doi: 10.1016/j.vaccine.2018.10.049] [Medline: 30389192]
41. Basch CH, Hillyer GC, Zagnit EA, Basch CE. YouTube coverage of COVID-19 vaccine development: implications for awareness and uptake. Hum Vaccin Immunother 2020 Nov 01;16(11):2582-2585. [doi: 10.1080/21645515.2020.1790280] [Medline: 32701403]
42. Kim J. The relationship of health beliefs with information sources and HPV vaccine acceptance among young adults in Korea. Int J Environ Res Public Health 2018 Apr 04;15(4):673 [FREE Full text] [doi: 10.3390/ijerph15040673] [Medline: 29617313]
43. Ford AJ, Alwan NA. Use of social networking sites and women's decision to receive vaccinations during pregnancy: a cross-sectional study in the UK. Vaccine 2018 Aug 23;36(35):5294-5303. [doi: 10.1016/j.vaccine.2018.07.022] [Medline: 30055969]
44. Luisi MLR. Who gives a “Like” about the HPV vaccine? Kansan parent/guardian perceptions and social media representations. KU ScholarWorks. 2018. URL: https://kuscholarworks.ku.edu/handle/1808/26037 [accessed 2021-03-74]
45. Hwang J, Shah DV. Health information sources, perceived vaccination benefits, and maintenance of childhood vaccination schedules. Health Commun 2019 Oct;34(11):1279-1288. [doi: 10.1080/10410236.2018.1481707] [Medline: 29869894]
46. Campbell H, Edwards A, Letley L, Bedford H, Ramsay M, Yarwood J. Changing attitudes to childhood immunisation in English parents. Vaccine 2017 May 19;35(22):2979-2985. [doi: 10.1016/j.vaccine.2017.03.089] [Medline: 28442229]
47. Wang Z, Fang Y, Ip M, Lau M, Lau JT. Facilitators and barriers to completing recommended doses of pneumococcal vaccination among community-living individuals aged ≥65 years in Hong Kong - a population-based study. Hum Vaccin Immunother 2021 Feb 01;17(2):527-536. [doi: 10.1080/21645515.2020.1776545] [Medline: 32614636]

48. Pan Y, Xin M, Zhang C, Dong W, Fang Y, Wu W, et al. Associations of mental health and personal preventive measure compliance with exposure to COVID-19 information during work resumption following the COVID-19 outbreak in China: cross-sectional survey study. J Med Internet Res 2020 Oct 08;22(10):e22596 [FREE Full text] [doi: 10.2196/22596] [Medline: 32936776]
49. Shenzhen Statistics Yearbook (2019). Statistics Bureau of Shenzhen Municipality. 2019. URL: http://tjj.sz.gov.cn/attachment/0/695/695422/7971762.pdf [accessed 2021-03-07]
50. Development and licensure of vaccines to prevent COVID-19: guidance for industry. US Food and Drug Administration. 2020. URL: http://www.fda.gov/media/139638/download [accessed 2020-12-09]
51. Wang Z, Wang J, Fang Y, Gross DL, Wong MC, Wong EL, et al. Parental acceptability of HPV vaccination for boys and girls aged 9-13 years in China - a population-based study. Vaccine 2018 May 03;36(19):2657-2665. [doi: 10.1016/j.vaccine.2018.03.057] [Medline: 29606519]
52. Wang Z, Fang Y, Dong W, Lau M, Mo PK. Illness representations on pneumonia and pneumococcal vaccination uptake among community-living Chinese people with high-risk conditions aged ≥65 years --- a population-based study. Hum Vaccin Immunother 2020 Sep 29:1-8. [doi: 10.1080/21645515.2020.1814653] [Medline: 32991245]
53. Zhang KC, Fang Y, Cao H, Chen H, Hu T, Chen YQ, et al. Parental acceptability of COVID-19 vaccination for children under the age of 18 years: cross-sectional online survey. JMIR Pediatr Parent 2020 Dec 30;3(2):e24827 [FREE Full text] [doi: 10.2196/24827] [Medline: 33326406]
54. A letter to enterprises in Shenzhen. Office of Shenzhen Headquarters for Prevention and Control of Pneumonia Epidemic Caused by Novel Coronavirus. 2020. URL: http://fao.sz.gov.cn/ydmh/xxgk/zyxw/202002/t20200205_18998951.htm [accessed 2020-12-08]
55. The State Council's joint prevention and control mechanism issued a notice calling for the effective strengthening of scientific epidemic prevention and control and orderly resumption of work and production of enterprises. Government of the People's Republic of China. 2020. URL: http://www.gov.cn/xinwen/2020-02/09/content_5476550.htm [accessed 2020-12-08]
56. Han J, Jia P, Huang Y, Gao B, Yu B, Yang S, et al. Association between social capital and mental health among older people living with HIV: the Sichuan Older HIV-Infected Cohort Study (SOHICS). BMC Public Health 2020 Apr 28;20(1):581 [FREE Full text] [doi: 10.1186/s12889-020-08705-6] [Medline: 32345273]
57. Lau CYK, Wang Z, Fang Y, Ip M, Wong KM, Chidgey A, et al. Prevalence of and factors associated with behavioral intention to take up home-based HIV self-testing among male clients of female sex workers in China - an application of the Theory of Planned Behavior. AIDS Care 2020 Jun 28:1-10. [doi: 10.1080/09540121.2020.1785996] [Medline: 32597197]
58. McEachan RRC, Conner M, Taylor NJ, Lawton RJ. Prospective prediction of health-related behaviours with the theory of planned behaviour: a meta-analysis. Health Psychol Rev 2011 Sep;5(2):97-144. [doi: 10.1080/17437199.2010.521684]
59. Wang Z, Feng T, Lau JTF, Kim Y. Acceptability of voluntary medical male circumcision (VMMC) among male sexually transmitted diseases patients (MSTDP) in China. PLoS One 2016;11(2):e0149801 [FREE Full text] [doi: 10.1371/journal.pone.0149801] [Medline: 26905739]
60. Pazol K, Robbins CL, Black LI, Ahrens KA, Daniels K, Chandra A, et al. Receipt of selected preventive health services for women and men of reproductive age - United States, 2011-2013. MMWR Surveill Summ 2017 Oct 27;66(20):1-31 [FREE Full text] [doi: 10.15585/mmwr.ss6620a1] [Medline: 29073129]
61. Liu B, Chen R, Zhao M, Zhang X, Wang J, Gao L, et al. Vaccine confidence in China after the Changsheng vaccine incident: a cross-sectional study. BMC Public Health 2019 Nov 27;19(1):1564 [FREE Full text] [doi: 10.1186/s12889-019-7945-0] [Medline: 31771543]
62. Stewart GL, Courtright SH, Barrick MR. Peer-based control in self-managing teams: linking rational and normative influence with individual and group performance. J Appl Psychol 2012 Mar;97(2):435-447. [doi: 10.1037/a0025303] [Medline: 21895352]
63. Ajzen I. Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. J Appl Soc Psychol 2002;32(4):665-683. [doi: 10.1111/j.1559-1816.2002.tb00236.x]
64. Lau JTF, Yeung NCY, Choi KC, Cheng MYM, Tsui HY, Griffiths S. Acceptability of A/H1N1 vaccination during pandemic phase of influenza A/H1N1 in Hong Kong: population based cross sectional survey. BMJ 2009 Oct 27;339:b4164 [FREE Full text] [doi: 10.1136/bmj.b4164] [Medline: 19861377]
65. Lau JT, Yeung NC, Choi K, Cheng MY, Tsui H, Griffiths S. Factors in association with acceptability of A/H1N1 vaccination during the influenza A/H1N1 pandemic phase in the Hong Kong general population. Vaccine 2010 Jun 23;28(29):4632-4637 [FREE Full text] [doi: 10.1016/j.vaccine.2010.04.076] [Medline: 20457289]

Abbreviations

AOR: adjusted odds ratio
CDC: Centre for Disease Control and Prevention
HPV: human papillomavirus
OR: odds ratio
QR: Quick Response

TPB: theory of planned behavior

Edited by C Basch; submitted 30.09.20; peer-reviewed by S Tomczyk, M Fahimi, S Teitler-Regev; comments to author 02.12.20; revised version received 16.12.20; accepted 26.02.21; published 09.03.21.

Please cite as:
Zhang KC, Fang Y, Cao H, Chen H, Hu T, Chen Y, Zhou X, Wang Z
Behavioral Intention to Receive a COVID-19 Vaccination Among Chinese Factory Workers: Cross-sectional Online Survey
J Med Internet Res 2021;23(3):e24673
URL: https://www.jmir.org/2021/3/e24673
doi:10.2196/24673
PMID:33646966

Review

Digital Health Solutions to Control the COVID-19 Pandemic in Countries With High Disease Prevalence: Literature Review

Sharareh R Niakan Kalhori[1], PhD; Kambiz Bahaadinbeigy[2], PhD; Kolsoum Deldar[3], PhD; Marsa Gholamzadeh[1], MSc; Sadrieh Hajesmaeel-Gohari[4], PhD; Seyed Mohammad Ayyoubzadeh[1], PhD

[1]Department of Health Information Management, Tehran University of Medical Sciences, Tehran, Iran

[2]Modeling in Health Research Center, Institute for Future Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

[3]School of Paramedicine, Shahroud University of Medical Sciences, Shahroud, Iran

[4]Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

Corresponding Author:
Seyed Mohammad Ayyoubzadeh, PhD
Department of Health Information Management
Tehran University of Medical Sciences
3rd Floor, School of Allied Medical Sciences
No #17, Farredanesh Alley, Ghods St, Enghelab Ave
Tehran
Iran
Phone: 98 2188983025
Email: s.m.ayyoubzadeh@gmail.com

Abstract

Background: COVID-19, the disease caused by the novel coronavirus SARS-CoV-2, has become a global pandemic, affecting most countries worldwide. Digital health information technologies can be applied in three aspects, namely digital patients, digital devices, and digital clinics, and could be useful in fighting the COVID-19 pandemic.

Objective: Recent reviews have examined the role of digital health in controlling COVID-19 to identify the potential of digital health interventions to fight the disease. However, this study aims to review and analyze the digital technology that is being applied to control the COVID-19 pandemic in the 10 countries with the highest prevalence of the disease.

Methods: For this review, the Google Scholar, PubMed, Web of Science, and Scopus databases were searched in August 2020 to retrieve publications from December 2019 to March 15, 2020. Furthermore, the Google search engine was used to identify additional applications of digital health for COVID-19 pandemic control.

Results: We included 32 papers in this review that reported 37 digital health applications for COVID-19 control. The most common digital health projects to address COVID-19 were telemedicine visits (11/37, 30%). Digital learning packages for informing people about the disease, geographic information systems and quick response code applications for real-time case tracking, and cloud- or mobile-based systems for self-care and patient tracking were in the second rank of digital tool applications (all 7/37, 19%). The projects were deployed in various European countries and in the United States, Australia, and China.

Conclusions: Considering the potential of available information technologies worldwide in the 21st century, particularly in developed countries, it appears that more digital health products with a higher level of intelligence capability remain to be applied for the management of pandemics and health-related crises.

(J Med Internet Res 2021;23(3):e19473) doi:10.2196/19473

KEYWORDS

COVID-19; digital health; information technology; telemedicine; electronic health

Introduction

The novel disease COVID-19, caused by the novel coronavirus SARS-CoV-2, was originally recognized in December 2019 as a case of pneumonia in Wuhan, China; it has since become a global pandemic, affecting most countries worldwide [1]. On March 11, the World Health Organization announced the outbreak of a pandemic and asked for coordinated mechanisms to support readiness and rapid response to the infection across the world's health sectors [2]. As the incidence of COVID-19

continues to rise, health care systems are rapidly facing growing clinical demands [3]. Operational management of a pandemic in the era of modern medicine requires novel technologies, such as digital health, that can support the management of COVID-19 cases in different stages [4]. Digital health as an application of information technology has already been used to improve health care organizations; for example, the National Health Service (NHS) in the United Kingdom has established the NHS Digital information center [5]. Digital health is defined as information technologies that can be applied in three aspects: digital patients, digital devices, and digital clinics. A digital patient is a patient who uses and engages with mobile health (mHealth) devices to change and sustain their behavior, including technologies such as telemedicine, patient self-measurements, and digital retention. Digital devices help solve clinical problems and include smartphone-connected rhythm monitoring devices, wireless and wearable devices, and implantable and ingestible sensors. The digital clinic aspect focuses on generating mHealth data, analyzing it so that it is clinically meaningful, and integrating it within clinical workflows. Aspects of digital clinics include precision-based mHealth and n-of-1 designs, population-based mHealth interventions in resource-limited areas, and mHealth regulation and integration [6].

During the COVID-19 pandemic, digital health–based tools may support organizations and societies more efficiently. They are useful for instant, widespread distribution of information, real-time transmission tracking, virtual venue creation for meetings and official day-to-day operations, and telemedicine visits for patients [7-12]. Such applications during the COVID-19 pandemic have been reported in several publications [13-15]. During the recent months of the COVID-19 outbreak, as countries and their responsible organizations such as health ministries and other officials have focused on controlling the pandemic, many supportive and reliable informatics infrastructures have been developed [12]. These infrastructures were applied in practice to prepare to manage an exponential increase in patients with COVID-19. Various digital health strategies have been used for disease control in different countries. A study conducted by Calton et al [16] provided some tips for applying telemedicine as a means to reduce the transmission of COVID-19. A study conducted by Moazzami et al [17] focused on employing telemedicine to prevent disease among health care providers. A study conducted by Keesara et al [18] referred to the capabilities and potential of digital health to fight COVID-19. However, they reviewed digital health–related solutions in general to address how this technology can support health care systems through introducing various strategic roles, such as surveillance, screening, triage, diagnosis, and monitoring, and contact tracing; no data regarding the use of this approach in practice for fighting COVID-19 were provided [19]. Fagherazzi et al [20] emphasized that the great potential of digital technology for COVID-19 control should be considered at the top level of health systems; they also discussed the challenges that policy makers may face in controlling the crisis using digital solutions. Furthermore, in a macro vision, they revealed the required societal and environmental restructuring required for successfully applying digital health technology to control COVID-19, including the health care system, government, public, industry, environment, and energy [15]. These reviews depict a general image regarding the requirement of digital system use and their applications worldwide [21], with no focus on any specific application in a specific country or region. Although these studies have shed light on the topic of applying digital health solutions for COVID-19 control, there is a gap of deep understanding regarding the application of these technologies in countries where COVID-19 is highly prevalent.

Therefore, this study aimed to review and analyze applied information technology and digital health–related strategies to control the COVID-19 pandemic in the 10 countries with the highest prevalence of the disease.

Methods

In this review study, the databases of Google Scholar, PubMed, Web of Science, and Scopus were searched in August 2020 to retrieve publications from December 2019 to August 15, 2020. The combination of keywords for searching is shown below:

> *(“Corona virus” OR “COVID 19” OR “coronavirus”) AND (computer OR internet OR web* OR mobile OR smart OR email OR video confer* OR telecommunication OR ICT OR “information technology” OR ehealth OR telehealth OR mHealth OR telecare OR telehealth OR telemedicine OR telemonitoring OR digital OR wearable OR IoT OR cloud) AND (Italy OR Spain OR USA OR France OR UK OR Iran OR China OR Netherlands OR Germany OR Belgium)*

The inclusion criteria were publications that introduce digital health applications to manage and control COVID-19 in humans, and the exclusion criteria were non-English publications, publications with no abstract, research on data analysis and modeling for prediction of epidemiological parameters, letters to the editor, and review studies. Data were analyzed using descriptive methods. Qualitative analysis of the included studies was performed based on predefined categories. A summary of the reviewed articles is provided in Table 1. Several items were analyzed in each paper, including (1) publication month; (2) country (Italy, Spain, United States, France, United Kingdom, Iran, China, the Netherlands, Germany, and Belgium, as they were the countries where COVID-19 was most prevalent according to the Worldometer website [22]); (3) purpose of the study, including screening, prevention, diagnosis, treatment, and follow-up of cases (defined as follows: screening: no symptom + no contact with COVID-19 patients; prevention: no symptom + contact with COVID-19 patients with no symptoms; diagnosis: having disease symptoms; treatment of COVID-19 cases: decreasing symptoms dramatically; and follow-up: discharged cases with the fewest symptoms); (4) scope and territory (village, city, region/province, state, country, and international), (5) digital tools, including robots, the Internet of Things, videoconferencing, web-based systems, cloud-based systems, wearable devices, clinical decision support systems (CDSSs), intelligent systems, smartphones, mobile apps, telecommunication systems, websites, digital media, and digital quick response (QR) codes.

Table 1. Details of the reviewed papers that discussed the application of digital health tools to control the COVID-19 pandemic.

Author	Journal	Publication month (2020)	Country	Purpose	Scope and territory	Applied digital tools	Application of digital tools
Kamel and Geraghty [23]	*International Journal of Health Geographics*	March	China	Prevention	International	Web-based systems, mobile apps, GIS[a]	Widespread distribution of information and real-time tracking of transmission
Yang et al [24]	*Clinical Oral Investigations*	May	China	Treatment and follow-up	Country	Web-based systems, mobile apps	Telemedicine visits for patients
Meng et al [25]	*International Journal of Clinical Pharmacy*	April	China	Treatment	Region	Cloud-based systems, smartphones, telecommunication systems	Provision of pharmaceutical care activities to patients and physicians by pharmacists
Ohannessian et al [26]	*JMIR Public Health and Surveillance*	February	France	Prevention	Country	Videoconferencing	Offering telemedicine visits for patients
Pan et al [27]	*Microbes and Infection*	February	China	Prevention	Country	Mobile apps	Widespread distribution of information and real-time tracking of transmission
Pan et al [28]	*Irish Journal of Medical Science*	March	China	Screening and prevention	City and country	Mobile apps	Real-time tracking of transmission
Sun et al [29]	*Annals of Intensive Care*	March	China	Treatment	State	Intelligent systems	Early warning systems and screening procedures for patients
Hernández-Garcia and Gimenez-Júlvez [30]	*JMIR Public Health and Surveillance*	April	Collaboration of the United States, Spain, Switzerland, the United Kingdom, Sweden, and Canada	Screening and prevention	International	Websites and digital media	Widespread distribution of information
Hua and Shaw [31]	*International Journal of Environmental Research and Public Health*	March	China	Screening, prevention, and follow-up	Region/ province	Web-based systems, smartphones, websites, digital media, digital QR[b] codes	Widespread distribution of information, real-time tracking of transmission, provision of information about "fake news" and rumors
Drew et al [32]	*Science*	May	United Kingdom, United States	Screening	International	Mobile app	Widespread distribution of information, real-time tracking of transmission
Franco et al [33]	*Global Spine Journal*	June	United States	Treatment	State	Videoconferencing, telephone	Offering telemedicine visits for patients
Gilbert et al [34]	*BMJ Open Quality*	May	United Kingdom	Prevention	City	Videoconferencing, telephone	Offering telemedicine visits for patients
Giudice et al [35]	*International Journal of Environmental Research and Public Health*	May	Italy	Follow-up	Region	Videoconferencing	Offering telemedicine visits for patients
Gong et al [36]	*Journal of Medical Internet Research*	April	China	Prevention	Country	Telecommunication system	Offering telemedicine visits for patients

Author	Journal	Publication month (2020)	Country	Purpose	Scope and territory	Applied digital tools	Application of digital tools
Gong et al [37]	*Journal of Medical Internet Research*	April	China	Screening	City	Cloud-based system, mobile app, CDSS[c]	Screening of cases and detection of patients
Goodman-Casanov et al [38]	*Journal of Medical Internet Research*	April	Spain	Prevention	Country	Telecommunication system	Widespread distribution of information, support for home care and patient self-care
Grange et al [39]	*Applied Clinical Informatics*	April	United States	Prevention, diagnosis, treatment, screening	State	Videoconferencing, CDSS, telecommunication	Offering telemedicine visits for patients
Grenda et al [40]	*Annals of Surgery*	August	United States	Diagnosis, treatment	City	Telecommunication, videoconferencing	Offering telemedicine visits for patients
Grossman et al [41]	*Neurology*	June	United States	Diagnosis, treatment	City	Smartphone, mobile apps	Offering telemedicine visits for patients
Hames et al [42]	*Journal of Psychotherapy Integration*	April	United States, Canada	Prevention	Country	Telecommunication system	Training
Hanna et al [43]	*Modern Pathology*	June	United States	Prevention, diagnosis	City	Telecommunication system	Diagnosis
Hom et al [44]	*Journal of Psychotherapy Integration*	April	United States	Prevention, treatment	City	Videoconferencing	Telemedicine visits for patients, training
Itamura et al [45]	*OTO Open*	April	United States	Prevention	Country	Videoconferencing	Telemedicine visits for patients
Judson et al [46]	*Journal of the American Medical Informatics Association*	June	United States	Prevention	State	Website	Screening of cases and detection of patients
Wu et al [47]	*European Respiratory Journal*	June	China, Italy, Belgium	Diagnosis	International	CDSS	Classification of patients in triage to find the best route
Wang et al [48]	*JMIR mHealth and uHealth*	June	China	Prevention	Country	Mobile app (WeChat)	Early tracing and quarantine of potential sources of infection
Timmers et al [49]	*JMIR mHealth and uHealth*	June	The Netherlands	Prevention	Country	Mobile app	Education, self-assessment, and symptom monitoring
Pepin et al [50]	*Journal of Medical Internet Research*	June	France	Prevention	International	Wearable devices and activity trackers	Definition of the level of quarantine
Rabuna et al [51]	*Telemedicine and e-Health*	June	Spain	Prevention	Rural area	TELEA digital web platform	Real-time tracking and monitoring of patients; follow-up of patients by telephone, videoconferencing, and email
Cheng et al [52]	Community Mental Health Journal	July	United States, Canada, Australia	Prevention	International	Mobile app	Peer-to-peer psychological support for Wuhan health care professionals at the front line of the crisis
Castaldi et al [53]	*Acta Biomedica*	July	Italy	Prevention	Region	Social media	Assessment of the dynamic burden of social anxiety through analysis of data from Facebook and Twitter

Author	Journal	Publication month (2020)	Country	Purpose	Scope and territory	Applied digital tools	Application of digital tools
Blake et al [54]	*International Journal of Environmental Research and Public Health*	July	United Kingdom	Prevention	Country	Digital learning package using agile methodology	Provision of psychologically safe spaces for staff through providing a three-step e-package with evidence-based guidance

[a]GIS: geographic information system.

[b]QR: quick response.

[c]CDSS: clinical decision support system.

Results

The search of scientific databases and manual searches retrieved 771 relevant articles. The titles and abstracts of all the retrieved publications were evaluated by two authors. Disagreements between the two evaluators were discussed and resolved by consensus. After removal of duplicates, 292 articles remained at this stage. Next, 260 publications were removed because they did not meet the inclusion criteria. Afterward, four authors independently reviewed the full text of the remaining publications (N=32). The reviewed papers were studied based on the variables shown in Table 1 and the different distributions discussed below.

For the purpose of this review, studies published from December 2019 to August 15, 2020, were reviewed. The survey identified 32 papers that demonstrated digital health applications to fight the COVID-19 pandemic. The distribution by publication month revealed that the publication of studies regarding digital health and COVID-19 began in February 2020, and the distribution of the 32 publications by month is February, 2 (6%); March, 5 (16%); April, 9 (28%); May, 3 (9%); June, 9 (28%); July, 3 (9%); and the first half of August, 1 (3%).

The projects of digital health application for COVID-19 control were deployed at different geographical levels, from international to rural. Six countries carried out six international projects, and the most common collaborations were among European countries, the United States, China, and Australia. The digital health projects at the international level mainly aimed to track real-time transmission and infected cases, define the level of quarantine, and enable peer-to-peer consultation to support care providers in other countries phytologically and scientifically. The studies of digital health projects for a given purpose in the 32 studies were most frequently conducted at the country level (n=10, 31%), and the other geographical levels were state (n=4, 13%), region (n=3, 9%), city (n=8, 25%), and rural (n=1, 3%). The United States was the country with the highest number of studies of digital health projects to fight COVID-19 (12/32, 38%), and these 12 studies varied the most in geographical scale, including international (n=3, 25%), state (n=3, 25%), country (n=2, 17%), and city (n=4, 13%) levels. The other studied countries ranked by the number of conducted studies were China (11/32, 34%); the United Kingdom (4/32, 13%); Canada, Spain, and Italy (3/32, 9%); Belgium and France (2/32, 6%); and the Netherlands (1/32, 3%).

To show the applied approaches of digital health for certain methods of COVID-19 control, the results were analyzed, and all the papers were categorized into six domains. These categories, their frequencies and percentages, and their applications for COVID-19 control are presented in Table 2. Some articles mentioned more than one approach to using digital health to control the COVID-19 pandemic.

Table 2. The frequency of digital health methods and their applications for COVID-19 pandemic control.

Domain number	Applied digital health solutions	COVID-19 control approaches	Digital health application projects (N=37), n (%)
1	Digital learning package, mobile apps, and web-based systems	Widespread distribution of information	7 (19)
2	GISs[a], QR[b] codes, and wearable devices	Real-time tracking of transmission, activity tracking, and quarantine-level analysis	7 (19)
3	Web-based systems and mobile apps, videoconferencing, and telephone	Telemedicine visit services and virtual venues for meetings	11 (30)
4	Cloud- and mobile-based systems	Self-care and patient monitoring, training, and diagnosis	7 (19)
5	Intelligent systems and CDSSs[c]	Early warning and detection, screening, and triage	4 (10)
6	Social media	Dynamic burden of the pandemic and analysis of its consequences	1 (3)

[a]GISs: geographic information systems.

[b]QR: quick response.

[c]CDSSs: clinical decision support systems.

According to the results, telemedicine visit services (11/37, 30%), especially in the United States (6/11, 54%), were the most commonly applied pandemic control approach. Using electronic methods to inform people about the disease, methods to prevent disease spread, and protection methods was the second-ranked approach, in addition to two other solutions of geographic information systems (GISs), QR codes, and wearable devices for real-time transmission tracking as well as cloud-based and mobile app usage for patient monitoring and self-care at home (all 7/37, 19%). A few studies were identified regarding the application of intelligent systems and CDSSs (4/37, 10%) and social media data analysis (1/37, 3%) for screening and burden of disease analysis purposes.

Discussion

Principal Findings

The COVID-19 pandemic has spread worldwide, costing lives and bringing upheaval and change to societies and economies. Although the global scientific community is racing to discover effective vaccines and therapeutics, the most essential defense remains public health measures such as personal hygiene and mass physical distancing. To successfully implement these two main measures, digital health and information technologies have emerged to support health systems, and they offer opportunities to reshape current health care systems. The aim of this study was to review the most significant digital health tools applied to fight COVID-19 in the 10 countries that have been most affected by the disease. These tools help governments and people to engage in strategies to control the COVID-19 pandemic through addressing the most urgent needs, including immediate outbreak response and impact mitigation. In China, which is the first country affected by the virus [55] and the most populous country globally, many researchers have worked on multiple aspects of SARS-CoV-2; it is the second most frequent origin country of the included studies. The burden of SARS-CoV-2 could be massive in populous countries; thus, these studies are worthy of investment in these countries. Studies that reported the development of models to predict epidemiological indicators were ignored, as they have not yet yielded any digital tools and require further development [56-59].

Distributing widespread information and tracking real-time transmission were the two most frequent goals of the studies. The former may originate from the importance of prevention in pandemic diseases as well as the simplest task of using information systems. The latter may be a focus in the literature because of the knowledge obtained from the previous experience of epidemics such as influenza and Zika virus [60-62]. Additionally, telemedicine visits for patients may be beneficial for populations because screening and follow-up of patients can be performed while maintaining social distancing in the population [63]. It appears that investigating the infrastructures needed for this technology could have great potential to mitigate these types of crises. In addition to the whole populations that can benefit from digital health technologies, more attention should be paid to interventions for travelers, as they can spread SARS-CoV-2 to other locations and even globally [64].

It has been shown that cell phones can be beneficial for health care [65]; due to their high influence among global populations, these tools are well suited for widespread distribution of information to these populations. Mobile apps are also used for tracking real-time transmission of SARS-CoV-2. Other potentially useful digital health tools are web-based apps and websites; these tools can also distribute information and track transmission. Videoconferencing and telecommunication also appear to be useful barriers to the spread of COVID-19 by enabling social distancing. Moreover, other industries may use teleservices to prevent the dissemination of disease.

Due to time limitations and the different times of onset of the epidemic in different countries, several digital health tools are not included in this paper. These tools have been reported in the news and other resources, and it may be valuable to discuss them as learned lessons for other countries fighting COVID-19. Therefore, we will review the digital health interventions in the different countries based on their available facilities and other requirements.

China

In China, multiple approaches are being used to manage the COVID-19 pandemic, ranging from web-based and mobile-based systems to cloud-based systems, CDSSs, and intelligent systems. The total number of cases of COVID-19 in this country showed a slight increase after March 1, 2020, based on the data in [66]. However, this decrease in COVID-19 cases was affected by multiple factors, and the effect of eHealth tools on the decrease should be evaluated. China has widely applied an eHealth app named Health Code to indicate a person's health status in the past day [67].

China has established a plan to spend approximately US $1.4 trillion on digital infrastructure. This infrastructure upgrade program includes developing 5G networks, industrial internet, data centers, and artificial intelligence [68], which could improve the country's capability to fight pandemics.

Italy and Spain

In contrast to Spain, Italy ranks among the four least advanced European countries in the Digital Economy and Society Index published by the European Commission [69], and approximately half the population of Italy has insufficient digital literacy [70]. The adoption of technology to prevent and manage the COVID-19 pandemic is unremarkable in these two countries. However, a coronavirus-tracking app was developed in Spain [71]. The statistics of total COVID-19 cases showed a dramatic increase after March 1, 2020, in both countries [66]. It appears that these countries should invest more in technologies to manage the pandemic.

United States

The US government launched a portal [72] for the public that contains information on how to prevent and manage COVID-19. Moreover, the US Centers for Disease Control and Prevention website [73] contains more detailed medical information on the spreading mechanism, symptoms, prevention, and treatment of COVID-19.

France and Belgium

The French app StopCovid was developed to trace infected people to control the spread of SARS-CoV-2. Privacy concerns arose regarding adoption of the app. Belgium has announced that a similar app adoption was canceled due to these issues [74].

United Kingdom

The NHS in the United Kingdom works on nine main areas to digitally respond to the pandemic: provide digital channels for citizen guidance and triage; enable remote and collaborative care with systems and data; deliver digital services for NHS Test and Trace; identify and protect vulnerable citizens; support planning with data, analysis, and dashboards; get data and insights to research communities; support clinical trials; provide secure infrastructure and support additional capacity; and plan for recovery, restarting services, and new needs. The government has categorized initiatives in these areas [75].

Iran

Although our study did not include any papers from this country, the Iranian Ministry of Health developed a national screening program website [76] to identify COVID-19 cases in the early stages.

The Netherlands

The Netherlands is one of the leading countries in Europe in digital health care and data. Approximately 90% of the population has digital records, and the Dutch government has invested over 400 million euros (US $482,980,000) in digital health. Hospitals in the Netherlands have signed up for a COVID-19 web-based portal for sharing patient information. Video consultation was provided by more than 8000 health care providers [77].

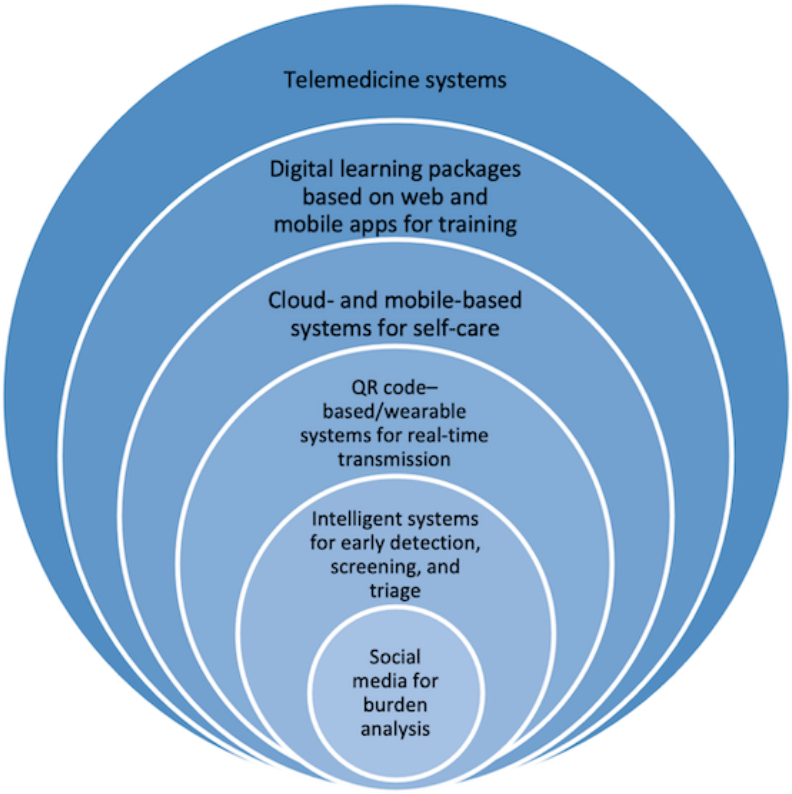
Germany

The Health Innovation Hub, established by Germany's Ministry of Health, has published a list of trusted telemedicine applications. The services provided by these apps include remote consultation, risk assessment, and telemedicine services. Before 2018, the country did not allow remote consultations.

The German parliament passed the Digital Care Act, which acknowledged that digital health is crucial for fighting the COVID-19 pandemic [78].

Telemedicine systems are highly used in many countries. In European countries, tracking of patients was adopted due to its feasibility in smaller countries; also, home care and self-care receive a relatively large amount of focus in these countries. Intelligent systems, CDSSs, and intelligent triage systems are not well adopted due to the need to supply them with data. These data are being gathered worldwide. Furthermore, analysis of social health data could be interesting, although little research has been done in this regard. Figure 1 shows the extent of the technologies developed for fighting the COVID-19 pandemic in the literature.

Figure 1. Technologies currently being applied to address the COVID-19 pandemic. QR: quick response.



Overall, in the studied countries, after the alarm was raised regarding the pandemic, implementation of eHealth strategies began immediately. mHealth solutions and large-scale deployment of virtual consultations were launched. Data analysis approaches are being applied to support decision makers, and websites and electronic training tools are being used to improve patients' protective behaviors. Although this study presents digital tools that are being applied for pandemic control in

general, it lacks evaluation of the exact outcomes of using these digital health tools; thus, further studies are needed to evaluate the effects and outcomes of using digital health tools. This study could help health policy makers make decisions regarding the investment of these tools to control COVID-19.

Conclusion

This study reviewed the digital health tools to fight COVID-19 that have been reported in the 10 countries in which the disease is most prevalent. Although there is no equal strategy to apply digital health tools across the affected countries for pandemic control, these tools are among the primary policies that governmental and private companies have considered for disease control. The United States has developed the most technologies to fight the pandemic. Furthermore, China, the first country that was affected by COVID-19, has applied a great number of digital tools, such as epidemiological indicators, analysis platforms, drones, robots, mobile apps, training websites and educational media, videoconferencing, smart infection detectors, intelligent patient tracers, and telemedicine systems. Having considered the potential of available information technologies worldwide in the 21st century, particularly in developed countries, it appears that more digital health products, especially intelligent products, remain to be created and applied for the management of viral infections and other health crises.

Conflicts of Interest

None declared.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733. [doi: 10.1056/nejmoa2001017]
2. Responding to community spread of COVID-19: interim guidance. World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/331421/WHO-COVID-19-Community_Transmission-2020.1-eng.pdf?sequence=1&isAllowed=y [accessed 2021-03-04]
3. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720. [doi: 10.1056/nejmoa2002032]
4. Atreja A, Gordon SM, Pollock DA, Olmsted RN, Brennan PJ, Healthcare Infection Control Practices Advisory Committee. Opportunities and challenges in utilizing electronic health records for infection surveillance, prevention, and control. Am J Infect Control 2008 Apr;36(3 Suppl):S37-S46 [FREE Full text] [doi: 10.1016/j.ajic.2008.01.002] [Medline: 18374211]
5. NHS Digital. URL: https://digital.nhs.uk/ [accessed 2021-03-06]
6. Bhavnani SP, Narula J, Sengupta PP. Mobile technology and the digitization of healthcare. Eur Heart J 2016 May 07;37(18):1428-1438 [FREE Full text] [doi: 10.1093/eurheartj/ehv770] [Medline: 26873093]
7. Joshi AU, Randolph FT, Chang AM, Slovis BH, Rising KL, Sabonjian M, et al. Impact of emergency department tele-intake on left without being seen and throughput metrics. Acad Emerg Med 2020 Feb 26;27(2):139-147 [FREE Full text] [doi: 10.1111/acem.13890] [Medline: 31733003]
8. Langabeer J, Gonzalez M, Alqusairi D, Champagne-Langabeer T, Jackson A, Mikhail J, et al. Telehealth-enabled emergency medical services program reduces ambulance transport to urban emergency departments. West J Emerg Med 2016 Nov;17(6):713-720 [FREE Full text] [doi: 10.5811/westjem.2016.8.30660] [Medline: 27833678]
9. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/nejmp2003539]
10. Lurie N, Carr BG. The role of telehealth in the medical response to disasters. JAMA Intern Med 2018 Jun 01;178(6):745-746. [doi: 10.1001/jamainternmed.2018.1314] [Medline: 29710200]
11. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
12. Reeves J, Hollandsworth H, Torriani F, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. J Am Med Inform Assoc 2020 Jun 01;27(6):853-859 [FREE Full text] [doi: 10.1093/jamia/ocaa037] [Medline: 32208481]
13. Scott BK, Miller GT, Fonda SJ, Yeaw RE, Gaudaen JC, Pavliscsak HH, et al. Advanced digital health technologies for COVID-19 and future emergencies. Telemed J E Health 2020 Oct 01;26(10):1226-1233. [doi: 10.1089/tmj.2020.0140] [Medline: 32456560]
14. Bayram M, Springer S, Garvey CK, Özdemir V. COVID-19 digital health innovation policy: a portal to alternative futures in the making. OMICS 2020 Aug 01;24(8):460-469. [doi: 10.1089/omi.2020.0089] [Medline: 32511054]
15. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
16. Calton B, Abedini N, Fratkin M. Telemedicine in the time of coronavirus. J Pain Symptom Manage 2020 Jul;60(1):e12-e14 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.03.019] [Medline: 32240756]

17. Moazzami B, Razavi-Khorasani N, Dooghaie Moghadam A, Farokhi E, Rezaei N. COVID-19 and telemedicine: Immediate action required for maintaining healthcare providers well-being. J Clin Virol 2020 May;126:104345 [FREE Full text] [doi: 10.1016/j.jcv.2020.104345] [Medline: 32278298]
18. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
19. Alwashmi MF. The use of digital health in the detection and management of COVID-19. Int J Environ Res Public Health 2020 Apr 23;17(8):2906 [FREE Full text] [doi: 10.3390/ijerph17082906] [Medline: 32340107]
20. Fagherazzi G, Goetzinger C, Rashid M, Aguayo G, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
21. Sarbadhikari S, Sarbadhikari S. The global experience of digital health interventions in COVID-19 management. Indian J Public Health 2020;64(6):117. [doi: 10.4103/ijph.ijph_457_20]
22. COVID-19 coronavirus pandemic. Worldometer. URL: https://www.worldometers.info/coronavirus/ [accessed 2021-03-04]
23. Kamel Boulos MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. Int J Health Geogr 2020 Mar 11;19(1):8 [FREE Full text] [doi: 10.1186/s12942-020-00202-8] [Medline: 32160889]
24. Yang Y, Zhou Y, Liu X, Tan J. Health services provision of 48 public tertiary dental hospitals during the COVID-19 epidemic in China. Clin Oral Investig 2020 May;24(5):1861-1864 [FREE Full text] [doi: 10.1007/s00784-020-03267-8] [Medline: 32246280]
25. Meng L, Qiu F, Sun S. Providing pharmacy services at cabin hospitals at the coronavirus epicenter in China. Int J Clin Pharm 2020 Apr 2;42(2):305-308 [FREE Full text] [doi: 10.1007/s11096-020-01020-5] [Medline: 32240484]
26. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 Pandemic: a call to action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
27. Pan X, Ojcius DM, Gao T, Li Z, Pan C, Pan C. Lessons learned from the 2019-nCoV epidemic on prevention of future infectious diseases. Microbes Infect 2020 Mar 27;22(2):86-91 [FREE Full text] [doi: 10.1016/j.micinf.2020.02.004] [Medline: 32088333]
28. Pan X. Application of personal-oriented digital technology in preventing transmission of COVID-19, China. Ir J Med Sci 2020 Nov 27;189(4):1145-1146 [FREE Full text] [doi: 10.1007/s11845-020-02215-5] [Medline: 32219674]
29. Sun Q, Qiu H, Huang M, Yang Y. Lower mortality of COVID-19 by early recognition and intervention: experience from Jiangsu Province. Ann Intensive Care 2020 Mar 18;10(1):33 [FREE Full text] [doi: 10.1186/s13613-020-00650-2] [Medline: 32189136]
30. Hernández-García I, Giménez-Júlvez T. Assessment of health information about COVID-19 prevention on the internet: infodemiological study. JMIR Public Health Surveill 2020 Apr 01;6(2):e18717 [FREE Full text] [doi: 10.2196/18717] [Medline: 32217507]
31. Hua J, Shaw R. Corona virus (COVID-19) "infodemic" and emerging issues through a data lens: the case of China. Int J Environ Res Public Health 2020 Mar 30;17(7):2309 [FREE Full text] [doi: 10.3390/ijerph17072309] [Medline: 32235433]
32. Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, COPE Consortium. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. Science 2020 Jun 19;368(6497):1362-1367 [FREE Full text] [doi: 10.1126/science.abc0473] [Medline: 32371477]
33. Franco D, Montenegro T, Gonzalez GA, Hines K, Mahtabfar A, Helgeson MD, et al. Telemedicine for the spine surgeon in the age of COVID-19: multicenter experiences of feasibility and implementation strategies. Global Spine J 2020 Jun 03:2192568220932168 [FREE Full text] [doi: 10.1177/2192568220932168] [Medline: 32677513]
34. Gilbert AW, Billany JCT, Adam R, Martin L, Tobin R, Bagdai S, et al. Rapid implementation of virtual clinics due to COVID-19: report and early evaluation of a quality improvement initiative. BMJ Open Qual 2020 May 21;9(2):e000985 [FREE Full text] [doi: 10.1136/bmjoq-2020-000985] [Medline: 32439740]
35. Giudice A, Barone S, Muraca D, Averta F, Diodati F, Antonelli A, et al. Can teledentistry improve the monitoring of patients during the Covid-19 dissemination? A descriptive pilot study. Int J Environ Res Public Health 2020 May 13;17(10):3399 [FREE Full text] [doi: 10.3390/ijerph17103399] [Medline: 32414126]
36. Gong K, Xu Z, Cai Z, Chen Y, Wang Z. Internet hospitals help prevent and control the epidemic of COVID-19 in China: multicenter user profiling Study. J Med Internet Res 2020 Apr 14;22(4):e18908 [FREE Full text] [doi: 10.2196/18908] [Medline: 32250962]
37. Gong M, Liu L, Sun X, Yang Y, Wang S, Zhu H. Cloud-based system for effective surveillance and control of COVID-19: useful experiences from Hubei, China. J Med Internet Res 2020 Apr 22;22(4):e18948 [FREE Full text] [doi: 10.2196/18948] [Medline: 32287040]
38. Goodman-Casanova JM, Dura-Perez E, Guzman-Parra J, Cuesta-Vargas A, Mayoral-Cleries F. Telehealth home support during COVID-19 confinement for community-dwelling older adults with mild cognitive impairment or mild dementia: survey study. J Med Internet Res 2020 May 22;22(5):e19434 [FREE Full text] [doi: 10.2196/19434] [Medline: 32401215]

39. Grange ES, Neil EJ, Stoffel M, Singh AP, Tseng E, Resco-Summers K, et al. Responding to COVID-19: The UW Medicine Information Technology Services Experience. Appl Clin Inform 2020 Mar 08;11(2):265-275 [FREE Full text] [doi: 10.1055/s-0040-1709715] [Medline: 32268390]
40. Grenda T, Whang S, Evans N. Transitioning a surgery practice to telehealth during COVID-19. Ann Surg 2020 Aug;272(2):e168-e169 [FREE Full text] [doi: 10.1097/SLA.0000000000004008] [Medline: 32675529]
41. Grossman SN, Han SC, Balcer LJ, Kurzweil A, Weinberg H, Galetta SL, et al. Rapid implementation of virtual neurology in response to the COVID-19 pandemic. Neurology 2020 Jun 16;94(24):1077-1087. [doi: 10.1212/WNL.0000000000009677] [Medline: 32358217]
42. Hames JL, Bell DJ, Perez-Lima LM, Holm-Denoma JM, Rooney T, Charles NE, et al. Navigating uncharted waters: considerations for training clinics in the rapid transition to telepsychology and telesupervision during COVID-19. J Psychother Integr 2020 Jun;30(2):348-365. [doi: 10.1037/int0000224]
43. Hanna MG, Reuter VE, Ardon O, Kim D, Sirintrapun SJ, Schüffler PJ, et al. Validation of a digital pathology system including remote review during the COVID-19 pandemic. Mod Pathol 2020 Nov 22;33(11):2115-2127 [FREE Full text] [doi: 10.1038/s41379-020-0601-5] [Medline: 32572154]
44. Hom MA, Weiss RB, Millman ZB, Christensen K, Lewis EJ, Cho S, et al. Development of a virtual partial hospital program for an acute psychiatric population: Lessons learned and future directions for telepsychotherapy. J Psychother Integr 2020 Jun;30(2):366-382. [doi: 10.1037/int0000212]
45. Itamura K, Rimell FL, Illing EA, Higgins TS, Ting JY, Lee MK, et al. Assessment of patient experiences in otolaryngology virtual visits during the COVID-19 pandemic. OTO Open 2020 Jun 08;4(2):2473974X2093357. [doi: 10.1177/2473974x20933573]
46. Judson T, Odisho A, Neinstein A, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):860-866 [FREE Full text] [doi: 10.1093/jamia/ocaa051] [Medline: 32267928]
47. Wu G, Yang P, Xie Y, Woodruff HC, Rao X, Guiot J, et al. Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicentre study. Eur Respir J 2020 Aug 02;56(2):2001104 [FREE Full text] [doi: 10.1183/13993003.01104-2020] [Medline: 32616597]
48. Wang S, Ding S, Xiong L. A new system for surveillance and digital contact tracing for COVID-19: spatiotemporal reporting over network and GPS. JMIR mHealth uHealth 2020 Jun 10;8(6):e19457 [FREE Full text] [doi: 10.2196/19457] [Medline: 32499212]
49. Timmers T, Janssen L, Stohr J, Murk J, Berrevoets MAH. Using eHealth to support COVID-19 education, self-assessment, and symptom monitoring in the Netherlands: observational study. JMIR mHealth uHealth 2020 Jun 23;8(6):e19822 [FREE Full text] [doi: 10.2196/19822] [Medline: 32516750]
50. Pépin JL, Bruno RM, Yang R, Vercamer V, Jouhaud P, Escourrou P, et al. Wearable activity trackers for monitoring adherence to home confinement during the COVID-19 pandemic worldwide: data aggregation and analysis. J Med Internet Res 2020 Jun 19;22(6):e19787 [FREE Full text] [doi: 10.2196/19787] [Medline: 32501803]
51. Rabuñal R, Suarez-Gil R, Golpe R, Martínez-García M, Gómez-Méndez R, Romay-Lema E, et al. Usefulness of a telemedicine tool TELEA in the management of the COVID-19 pandemic. Telemed J E Health 2020 Nov 01;26(11):1332-1335. [doi: 10.1089/tmj.2020.0144] [Medline: 32501747]
52. Cheng P, Xia G, Pang P, Wu B, Jiang W, Li Y, et al. COVID-19 epidemic peer support and crisis intervention via social media. Community Ment Health J 2020 Jul 6;56(5):786-792 [FREE Full text] [doi: 10.1007/s10597-020-00624-5] [Medline: 32378126]
53. Castaldi S, Maffeo M, Rivieccio B, Zignani M, Manzi G, Nicolussi F, et al. Monitoring emergency calls and social networks for COVID-19 surveillance. To learn for the future: the outbreak experience of the Lombardia region in Italy. Acta Biomed 2020 Jul 20;91(9-S):29-33. [doi: 10.23750/abm.v91i9-S.10038] [Medline: 32701914]
54. Blake H, Bermingham F, Johnson G, Tabner A. Mitigating the psychological impact of COVID-19 on healthcare workers: a digital learning package. Int J Environ Res Public Health 2020 Apr 26;17(9):2997 [FREE Full text] [doi: 10.3390/ijerph17092997] [Medline: 32357424]
55. Guo Y, Cao Q, Hong Z, Tan Y, Chen S, Jin H, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. Mil Med Res 2020 Mar 13;7(1):11 [FREE Full text] [doi: 10.1186/s40779-020-00240-0] [Medline: 32169119]
56. Ayyoubzadeh S, Ayyoubzadeh S, Zahedi H, Ahmadi M, R Niakan Kalhori S. Predicting COVID-19 incidence through analysis of Google Trends data in Iran: data mining and deep learning pilot study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18828 [FREE Full text] [doi: 10.2196/18828] [Medline: 32234709]
57. Qiang X, Xu P, Fang G, Liu W, Kou Z. Using the spike protein feature to predict infection risk and monitor the evolutionary dynamic of coronavirus. Infect Dis Poverty 2020 Mar 25;9(1):33 [FREE Full text] [doi: 10.1186/s40249-020-00649-8] [Medline: 32209118]
58. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis 2020 Jun;20(6):669-677. [doi: 10.1016/s1473-3099(20)30243-7]

59. Wu J, Leung K, Leung G. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020 Feb;395(10225):689-697. [doi: 10.1016/s0140-6736(20)30260-9]
60. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining search, social media, and traditional data sources to improve influenza surveillance. PLoS Comput Biol 2015 Oct 29;11(10):e1004513 [FREE Full text] [doi: 10.1371/journal.pcbi.1004513] [Medline: 26513245]
61. McGough SF, Brownstein JS, Hawkins JB, Santillana M. Forecasting Zika incidence in the 2016 Latin America outbreak combining traditional disease surveillance with search, social media, and news report data. PLoS Negl Trop Dis 2017 Jan 13;11(1):e0005295. [doi: 10.1371/journal.pntd.0005295] [Medline: 28085877]
62. Majumder MS, Santillana M, Mekaru SR, McGinnis DP, Khan K, Brownstein JS. Utilizing nontraditional data sources for near real-time estimation of transmission dynamics during the 2015-2016 Colombian Zika virus disease outbreak. JMIR Public Health Surveill 2016 Jun 01;2(1):e30 [FREE Full text] [doi: 10.2196/publichealth.5814] [Medline: 27251981]
63. Serper M, Cubell AW, Deleener ME, Casher TK, Rosenberg DJ, Whitebloom D, et al. Telemedicine in liver disease and beyond: can the COVID-19 crisis lead to action? Hepatology 2020 Aug 25;72(2):723-728 [FREE Full text] [doi: 10.1002/hep.31276] [Medline: 32275784]
64. Wilson ME. What goes on board aircraft? Passengers include Aedes, Anopheles, 2019-nCoV, dengue, Salmonella, Zika, et al. Travel Med Infect Dis 2020;33:101572 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101572] [Medline: 32035269]
65. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
66. Jibb LA, Cafazzo JA, Nathan PC, Seto E, Stevens BJ, Nguyen C, et al. Development of a mHealth real-time pain self-management app for adolescents with cancer: an iterative usability testing study. J Pediatr Oncol Nurs 2017 Apr 04;34(4):283-294. [doi: 10.1177/1043454217697022] [Medline: 28376666]
67. Kim H, Lee H, Park JI, Choi CH, Park S, Kim HJ, et al. Smartphone application for mechanical quality assurance of medical linear accelerators. Phys Med Biol 2017 Jun 07;62(11):N257-N270. [doi: 10.1088/1361-6560/aa67d5] [Medline: 28319042]
68. Wong D. How Can Foreign Technology Investors Benefit from China's New Infrastructure Plan? China Briefing. 2020 Aug 07. URL: https://www.china-briefing.com/news/how-foreign-technology-investors-benefit-from-chinas-new-infrastructure-plan/ [accessed 2021-03-04]
69. The Digital Economy and Society Index (DESI). European Commission. URL: https://ec.europa.eu/digital-single-market/en/digital-economy-and-society-index-desi [accessed 2021-03-08]
70. Guerrini F. How the coronavirus is forcing Italy to become a digital country, at last. Forbes. 2020 Mar 14. URL: https://www.forbes.com/sites/federicoguerrini/2020/03/14/how-the-coronavirus-is-forcing-italy-to-become-a-digital-country-at-last/?sh=649ae9376f75 [accessed 2021-03-04]
71. Vega G. Spain launches first phase of coronavirus-tracking app. El País. 2020 Jun 29. URL: https://english.elpais.com/society/2020-06-29/spain-launches-first-phase-of-coronavirus-tracking-app.html [accessed 2021-03-04]
72. Coronavirus (COVID-19). US Centers for Disease Control and Prevention and Federal Emergency Management Agency. URL: https://www.coronavirus.gov/ [accessed 2021-03-06]
73. COVID-19.: Centers for Disease Control and Prevention URL: https://www.cdc.gov/coronavirus/2019-ncov/index.html [accessed 2021-03-06]
74. Thompson R. StopCOVID: France's controversial tracing app ready by June. Euronews. 2020 May 05. URL: https://www.euronews.com/2020/04/29/coronavirus-french-mps-approve-covid-19-tracing-app-despite-privacy-concerns [accessed 2021-03-04]
75. NHS Digital COVID-19 Update. NHS Digital. 2020 Apr 30. URL: https://digital.nhs.uk/binaries/content/assets/website-assets/coronavirus/nhs-digital-covid-19-update-april-2020.pdf [accessed 2021-03-06]
76. New corona screening and care (COVID-19). Iran Ministry of Health and Medical Education. URL: https://salamat.gov.ir/ [accessed 2021-03-06]
77. How the Dutch are responding to coronavirus with digital healthcare. Invest in Holland. 2020 May 01. URL: https://investinholland.com/news/how-dutch-have-responded-digitally-corona-crisis/#:~:text=Pharos%20responded%20to%20this%20challenge,joint%2Dcooperation%20with%20other%20parties.&text=In%20the%20Netherlands%2C%20the%20Dutch,step%20up%20innovative%20eHealth%20solutions [accessed 2021-03-04]
78. How Germany leveraged digital health to combat COVID-19. The Medical Futurist. 2020 Apr 09. URL: https://medicalfuturist.com/how-germany-leveraged-digital-health-to-combat-covid-19/ [accessed 2021-03-04]

Abbreviations

CDSS: clinical decision support system
GIS: geographic information system
mHealth: mobile health
NHS: National Health Service

QR: quick response

Edited by G Eysenbach, T Sanchez; submitted 19.04.20; peer-reviewed by B Chaudhry, A Dormanesh, PG Tian, J Van den Eynde; comments to author 12.06.20; revised version received 27.06.20; accepted 31.08.20; published 10.03.21.

Please cite as:
R Niakan Kalhori S, Bahaadinbeigy K, Deldar K, Gholamzadeh M, Hajesmaeel-Gohari S, Ayyoubzadeh SM
Digital Health Solutions to Control the COVID-19 Pandemic in Countries With High Disease Prevalence: Literature Review
J Med Internet Res 2021;23(3):e19473
URL: https://www.jmir.org/2021/3/e19473
doi:10.2196/19473
PMID:33600344

Review

Evidence Synthesis of Digital Interventions to Mitigate the Negative Impact of the COVID-19 Pandemic on Public Mental Health: Rapid Meta-review

Christian Rauschenberg[1,2*], MSc; Anita Schick[1*], PhD, Dipl-Psych; Dusan Hirjak[3], MD; Andreas Seidler[4], MD, MPH; Isabell Paetzold[1], MSc; Christian Apfelbacher[5], PhD, Dr sc hum, MSc, MA; Steffi G Riedel-Heller[6], MD, MPH; Ulrich Reininghaus[1,7,8], PhD, MSc, Dipl-Psych

[1]Department of Public Mental Health, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[2]Department of Psychiatry and Neuropsychology, School of Mental Health and Neuroscience, Maastricht University, Maastricht, Netherlands

[3]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[4]Institute and Policlinic of Occupational and Social Medicine, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[5]Institute of Social Medicine and Health Systems Research, Otto von Guericke University Magdeburg, Magdeburg, Germany

[6]Institute of Social Medicine, Occupational Health and Public Health, University of Leipzig, Leipzig, Germany

[7]Centre for Epidemiology and Public Health, Health Service and Population Research Department, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[8]ESRC Centre for Society and Mental Health, King´s College London, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Christian Rauschenberg, MSc
Department of Public Mental Health
Central Institute of Mental Health
Medical Faculty Mannheim, Heidelberg University
J5
Mannheim, 68159
Germany
Phone: 49 62117031929
Email: christian.rauschenberg@zi-mannheim.de

Abstract

Background: Accumulating evidence suggests the COVID-19 pandemic has negative effects on public mental health. Digital interventions that have been developed and evaluated in recent years may be used to mitigate the negative consequences of the pandemic. However, evidence-based recommendations on the use of existing telemedicine and internet-based (eHealth) and app-based mobile health (mHealth) interventions are lacking.

Objective: The aim of this study was to investigate the theoretical and empirical base, user perspective, safety, effectiveness, and cost-effectiveness of digital interventions related to public mental health provision (ie, mental health promotion, prevention, and treatment of mental disorders) that may help to reduce the consequences of the COVID-19 pandemic.

Methods: A rapid meta-review was conducted. The MEDLINE, PsycINFO, and CENTRAL databases were searched on May 11, 2020. Study inclusion criteria were broad and considered systematic reviews and meta-analyses that investigated digital tools for health promotion, prevention, or treatment of mental health conditions and determinants likely affected by the COVID-19 pandemic.

Results: Overall, 815 peer-reviewed systematic reviews and meta-analyses were identified, of which 83 met the inclusion criteria. Our findings suggest that there is good evidence on the usability, safety, acceptance/satisfaction, and effectiveness of eHealth interventions. Evidence on mHealth apps is promising, especially if social components (eg, blended care) and strategies to promote adherence are incorporated. Although most digital interventions focus on the prevention or treatment of mental disorders, there is some evidence on mental health promotion. However, evidence on process quality, cost-effectiveness, and long-term effects is very limited.

Conclusions: There is evidence that digital interventions are particularly suited to mitigating psychosocial consequences at the population level. In times of physical distancing, quarantine, and restrictions on social contacts, decision makers should develop digital strategies for continued mental health care and invest time and efforts in the development and implementation of mental health promotion and prevention programs.

(J Med Internet Res 2021;23(3):e23365) doi:10.2196/23365

KEYWORDS

COVID-19; mHealth; eHealth; telemedicine; prevention; mental health promotion; intervention; digital mental health; digital intervention; public mental health

Introduction

Measures to prevent and control infections during the COVID-19 pandemic such as physical distancing, quarantine, and restrictions on social contacts can have a negative impact on public mental health [1]. This includes an increase in depression, anxiety, loneliness, and perceived stress [2] as well as in risk behaviors such as cannabis and alcohol use [3] in the population. In addition to the immediate effects of the infection control measures, further negative consequences for mental health are to be expected due to the more direct, deleterious effects of COVID-19 (eg, illness anxiety, contamination fears) as well as the economic downturn and recession [4]. Recently reported restrictions in access to, and continuity of, care for individuals with mental disorder caused by infection prevention and control measures in some countries are an additional cause for concern [3,5,6].

Digital interventions that do not require face-to-face contact may play an important role in improving public mental health at times of infection prevention and control measures. They can be broadly grouped as telemedicine and internet-based interventions (hereafter eHealth interventions) [7] and app-based mobile health (mHealth) interventions delivered using smartphones or other mobile devices [8]. These interventions provide a unique opportunity for delivering low-threshold, public mental health care tailored to individual needs and contexts in daily life, outside the clinic [9], even under the restrictive conditions of the COVID-19 pandemic. As smartphones are mostly in close proximity to users, and accessible whenever and wherever it is convenient, the use of mHealth apps in particular represents a powerful approach that allows for the real-time and real-world delivery of intervention components in individuals' daily lives.

Digital tools may help to mitigate negative psychosocial consequences most effectively if intervention strategies are not only targeted at vulnerable individuals in a clinically high-risk state or with a mental disorder but also at the population level. More specifically, following the seminal "population strategy" advocated by Rose [10], even a small shift in the population's mean level of mental health, which is continuously distributed in the population, may lead to a substantial reduction of the prevalence of mental health problems. If applied to the current pandemic, a scalable digital public mental health approach may contribute to lower rates of mental disorders by targeting important determinants and shifting the mean level of mental health in the population.

In order to minimize the negative impact of the COVID-19 pandemic on the mental health of the population, digital interventions can be used in the following areas of public mental health provision: primary prevention strategies, including (1) mental health promotion and literacy at the population level; (2) indicated, selective, or universal prevention targeting high-risk individuals, subpopulations, or the entire population, respectively, as well as secondary and tertiary prevention strategies, including (3) treatment and preventive services for people with mental disorders. Indeed, evidence from ad hoc surveys suggests that digital interventions for improving public mental health are urgently needed to address the psychosocial consequences of the COVID-19 pandemic [1-3,11,12]. For example, findings from the serial cross-sectional survey *German COVID-19 Snapshot Monitoring* (COSMO Germany [13]) suggest strong concerns about the economy, social inequalities, and the health care system as well as high levels of psychological distress in the adult general population, particularly among young people [14,15]. Another representative survey (Norstatpanel) found that a staggering 38% of youth met the criteria for moderate or severe mental health problems, even after the most restrictive infection control measures had been lifted [16]. Furthermore, the reported social isolation during the COVID-19 pandemic was associated with levels of psychological distress in a dose-response fashion [16]. Recent evidence also suggests a high subjective demand for digital mental health interventions in the general population and people with a mental disorder [17,18], which is matched with a high and rapidly growing number of mHealth apps available in major app stores, with the strongest growth having been noted for mHealth apps [19]. It has further been reported that the demand for mHealth apps has increased globally by 49% during the COVID-19 pandemic [20], with 73% of psychologically distressed and socially isolated youth in the Norstatpanel survey indicating the use of mHealth apps to be helpful in coping with the ongoing COVID-19 pandemic [16].

Taken together, based on the evidence presented, there is an urgent need for, and high potential in, using digital interventions to improve public mental health and mitigate the negative psychosocial impact of the COVID-19 pandemic. However, evidence-based recommendations for the use of digital interventions during public health crises, including this ongoing pandemic, is currently lacking. The present meta-review aimed to synthesize the available evidence on the theoretical and empirical base of interventions, quality assessments from the user perspective (ie, acceptability, usability, satisfaction), safety, effectiveness, and cost-effectiveness of digital interventions in

the area of public mental health provision (ie, mental health promotion and prevention of and treatment for mental disorder).

Methods

Overview

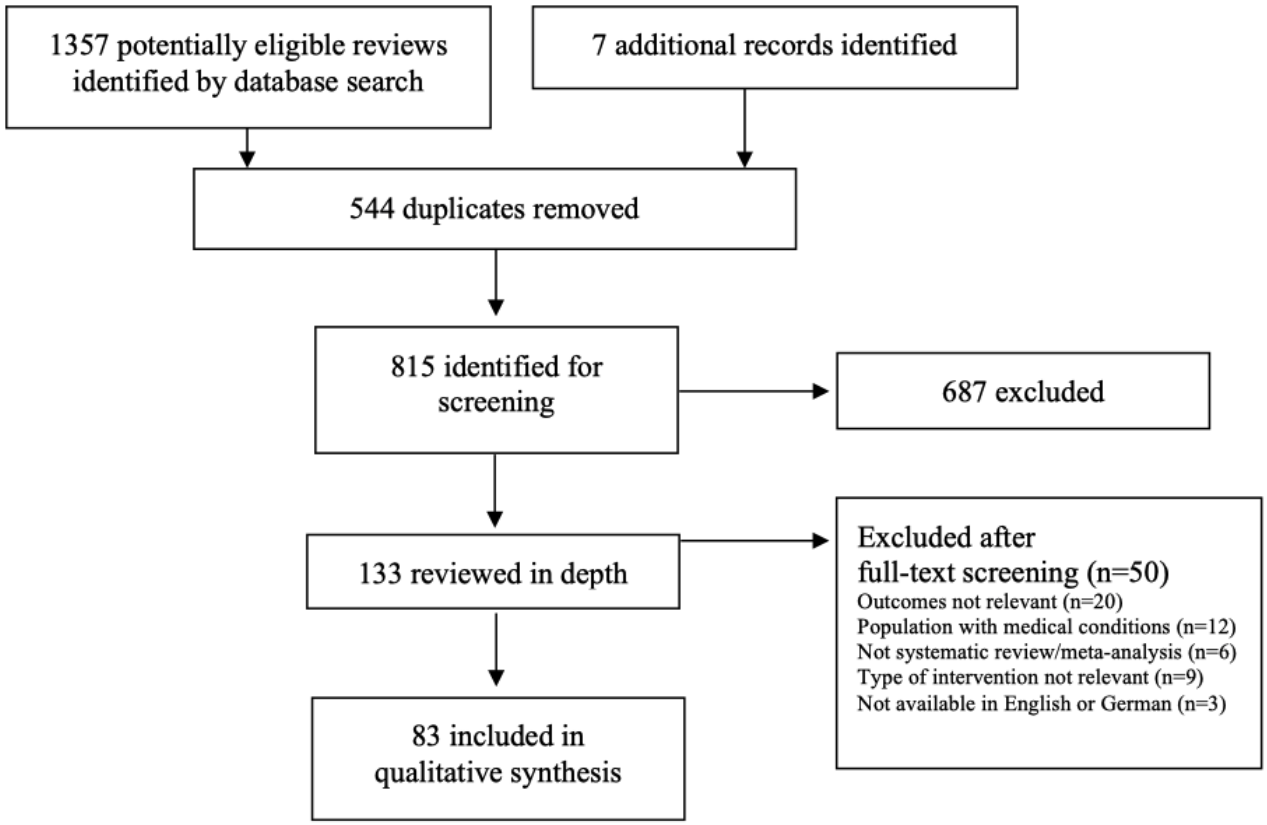
A rapid meta-review of systematic reviews on digital public mental health interventions was conducted. For this, PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses [21]) was used as a guideline for reporting findings. In line with the current state of the art in the development and evaluation of complex digital mental health interventions [8], the following criteria to review the available evidence were used: theoretical and evidence base, quality assessments from the user perspective (ie, acceptability, usability, satisfaction), safety, effectiveness, and cost-effectiveness.

Search Strategy and Selection Criteria

The MEDLINE, PsycINFO, and CENTRAL databases were searched for systematic reviews and meta-analyses published in the English and German languages from inception to April 2020. An extensive search of bibliographic databases was performed using queries that combined search terms on mental health, public mental health provision, digital eHealth/mHealth interventions (Multimedia Appendix 1), and high-quality reviews (ie, systematic review, meta-analysis) using logical operators. In doing so, database-specific queries were used to ensure semantic equivalence. The queries were launched on May 11, 2020, covering results until April 2020. The results were obtained, and duplicates were removed. References written in English and German were included. No other filters or restrictions were applied.

The search criteria were purposefully broad and considered systematic reviews and meta-analyses that investigated digital tools for health promotion, prevention, or treatment of mental health conditions and determinants likely affected by the COVID-19 pandemic (eg, depression, anxiety, psychosis, substance misuse, self-harm, well-being, quality of life, self-esteem, loneliness). Titles and abstracts were screened for inclusion by 1 reviewer (a research assistant). Studies were included if they were published in a peer-reviewed journal, contained original findings examining the theoretical and evidence base, quality from the user perspective (ie, acceptability, usability, satisfaction), safety, effectiveness, or cost-effectiveness of digital mHealth and eHealth interventions. Due to the rapid meta-review format of our study, conclusions drawn by the authors of the included systematic reviews were reported. The included articles had to be systematic reviews and/or meta-analyses that followed established reporting guidelines (eg, PRISMA [21]). Because of time constraints and the rapid meta-review format of this study, a second reviewer (CR) independently screened a randomly selected subset (40%) of identified studies. The references were categorized as "eligible," "query," and "not eligible." Inclusion and exclusion criteria were applied to references that were queried or eligible. Reviewers were blinded, and potential discrepancies in selection decisions were discussed with another member of the research team. A pilot screening of a randomly selected subset of identified studies (around 5%) was conducted to discuss decisions on categorizing studies at an early stage. As inclusion criteria were purposefully broad, discrepancies between the reviewers (CR and the research assistant) were very low. Full texts of potentially relevant articles were obtained, read, and assessed by 1 reviewer (CR), and data extraction was performed by 3 reviewers (CR and 2 research assistants under the close supervision of CR; see acknowledgments). Reviews and meta-analyses on preprint servers and gray literature were not included. The EndNote reference management software [22] was used to record reviewers' decisions, including reasons for exclusion. The study selection process was documented using the PRISMA flow diagram (Figure 1).

Figure 1. Study selection.



Results

General Findings

The search strategy of our meta-review on digital interventions yielded 815 peer-reviewed systematic reviews and meta-analyses (Figure 1). Of these, 83 references were included in the meta-review. Overall, 44 of the included reviews summarized findings on eHealth interventions and focused on interventions targeting depression (n=19), anxiety (n=22), problematic substance use (n=11), and eating disorders (n=2). Several reviews included interventions that targeted multiple mental health problems. In total, 16 reviews summarized findings on mHealth interventions and focused on depression (n=2), anxiety (n=1), problematic substance use (n=1), and eating disorder (n=1). Similarly, the majority of included reviews targeted various mental health domain. Furthermore, 23 of the included reviews jointly reported the effects of eHealth and mHealth interventions on various mental health outcomes (Multimedia Appendix 2). A complete summary of included reviews on eHealth, mHealth, and mixed interventions are shown in Multimedia Appendix 2, including findings on their theoretical and empirical base, user perspective, safety, effectiveness, and cost-effectiveness.

Theoretical and Empirical Base

For most eHealth and mHealth interventions in the area of mental health promotion and prevention as well as treatment of mental disorders, the theoretical and empirical base is explicitly mentioned in the description of interventions and are often informed by clinical guidelines and co-designed by service users and mental health professionals [23-26]. This primarily includes evidence-based procedures such as cognitive-behavioral therapy (CBT) or third-wave CBT [23,25,26]. In contrast to digital interventions developed by research groups, prominent mHealth apps available in major app stores do often not provide information on the theoretical or empirical base of their content [8,27-29]. Some mHealth apps may even be harmful and hinder healing processes (eg, asking users to do tasks that are too difficult to complete, presenting means for self-harm as well as lethal means, triggering unwanted distressing memories) [8,27,28].

Quality Assessments From the User Perspective

Evidence from the included systematic reviews suggests moderate to high levels of acceptance, feasibility, and user satisfaction with eHealth and mHealth interventions for mental health promotion and prevention [30,31] as well as for the treatment of mental health problems [32-41]. This applies, in particular, for interventions including social components [32,42], strategies to promote user adherence [33,43], symptom monitoring [44,45], or a blended-care approach [46].

In terms of safety, data sharing and data safety regulations, as well as aspects of eHealth/mHealth and clinical safety of interventions, were often not explicitly reported or systematically investigated in the identified systematic reviews [47-49] (Multimedia Appendix 2). The descriptions of many eHealth interventions do not make explicit reference to prevailing regulations and clinical guidelines [50]. Furthermore, there is evidence that mHealth apps available in major app stores use problematic data sharing and privacy practices (eg,

monetization of sensitive user data through analytics and advertising) [8,27,28].

Effectiveness of eHealth Interventions

There was good evidence on the effectiveness of telemedical and other eHealth interventions in the field of mental health promotion and prevention, as well as for the treatment of mental health conditions.

Mental Health Promotion and Prevention

There have been a number of systematic reviews that aimed to investigate the effectiveness of telemedical and eHealth interventions for mental health promotion and prevention. These interventions have primarily been shown to improve mental health [34], physical activity [34,35], well-being [36,37], stress [23,38], depression [23,36,38,51,52], anxiety [23,36,38,51,52], and alcohol [24,53-56] and cannabis use [57,58] in the general population in addition to dysfunctional cognition and self-esteem in at-risk populations [59,60]. Importantly, effectiveness has been demonstrated across differing age groups, including adults [24,54,59] and adolescents from the general population [34,52,56,61-63], and effect sizes mostly ranged from small to medium. However, evidence on the use of eHealth interventions for the elderly is scarce, although findings from the identified reviews indicated some evidence on the effectiveness of eHealth interventions for reducing social isolation and increasing social participation of people aged 65 years and older [64], which may be of particular interest in the context of the COVID-19 pandemic.

Treatment of Mental Health Conditions

There was also strong evidence on the effectiveness of telemedical and eHealth interventions in the provision of treatment and services for people with mental disorder. This included anxiety disorders [65-68], depression [60,61,65-67,69-73], substance abuse [54,74-76], eating disorders [77], and severe mental illness [78], with overall small to medium effect sizes, not only with regard to the reduction of relevant symptoms but also improvements in dysfunctional cognition [60], self-esteem [60], and quality of life [66]. Some of the identified studies have even reported medium to large effect sizes for cognitive-behavioral eHealth interventions that aimed to reduce symptoms of depression [79].

The effectiveness of telemedicine interventions that use videoconference tools or the telephone has also been well documented in depressive [80-83], anxiety [80,83-85], and psychotic disorders [86], with comparable effects for online group and individual therapy sessions [87,88], compared with conventional (offline) therapy sessions. Superior effectiveness was observed for interventions adopting a blended-care approach combining eHealth with conventional intervention components [46,54,71].

Overall, findings suggest that the evidence on long-term effects and noninferiority compared to conventional therapy and active control conditions remains limited [79,81,82,86,87]. There is also limited evidence on the impact of telemedical and eHealth interventions on underlying processes and mechanisms of change [89].

Effectiveness of mHealth Interventions

While there is some initial evidence on the effectiveness of mHealth interventions to improve physical activity [90-95], stress appraisal [96,97], depression [26,96-100], anxiety [25,26,96,97], and alcohol and substance use [55,96,98,101-103], with small to medium effect sizes in all areas of public mental health provision, the amount of research to investigate this issue remains, overall, limited [104-108]. Only a minority of mHealth interventions were found to use more advanced techniques (accelerometer, GPS) to inform the delivery of intervention components [25,89,92]. In addition, a substantial difference was found between mHealth apps available in major app stores, for which there is no or only very limited evidence on their effectiveness [29,108-111], and mHealth interventions developed by research groups. Similar to eHealth interventions, evidence on long-term effects and on underlying processes and mechanisms of action remains very limited.

Cost-effectiveness

There is some evidence on the cost-effectiveness of eHealth interventions for depression and anxiety in primary care settings when compared to care as usual and waiting list control conditions [51] as well as for a range of mental disorders when compared to conventional CBT [112,113]. However, as only a few systematic reviews have systematically investigated the cost-effectiveness of digital interventions to date, these findings should be interpreted with caution. While there is some evidence on the cost-effectiveness of mHealth interventions (eg, for digital monitoring and feedback in depression) from individual studies [18], evidence summarized at the level of systematic reviews is very limited.

Discussion

Principal Results

Evidence-based eHealth and mHealth interventions may play a central role in areas of public mental health provision (ie, mental health promotion, as well as prevention of and treatment for mental disorders) to mitigate the negative consequences of the COVID-19 pandemic. To date, however, evidence-based recommendations on existing digital interventions that have been developed and evaluated in recent years are lacking. This meta-review was the first to review the available evidence on the theoretical and empirical base, quality assessments from the user perspective (ie, acceptability, usability, satisfaction), safety, effectiveness, and cost-effectiveness of digital interventions in the area of public mental health provision, that is, mental health promotion at the population level, indicated, selective, or universal prevention targeting high-risk individuals, subpopulations, or the entire population as well as treatment and services for people with mental disorders.

First, there was robust evidence on the effectiveness of telemedical eHealth interventions and initial evidence on the effectiveness of mHealth interventions in relation to mental health outcomes likely affected by the COVID-19 pandemic (eg, anxiety, depression), especially if interventions are informed by clinical guidelines and co-designed by service users and

mental health professionals. Second, effectiveness, acceptability, feasibility, and user satisfaction have been described to be particularly high if digital interventions are embedded in a therapeutic context and include some form of social interaction with a mental health professional (blended-care approach). Third, some of the included systematic reviews and meta-analyses suggest noninferiority of effectiveness for some eHealth interventions as compared to traditional face-to-face therapy, but further replication is needed before firm conclusions can be drawn. Thus, in order to exclude the risk of infection in the current public health crisis, clinicians and other health professionals may consider combining differing types of digital interventions (eg, counseling or psychotherapy using videoconference software augmented by a smartphone-based mHealth app) as this approach may be particularly promising given the current evidence base and reflects a novel digital version of the blended-care approach. However, more research is needed to investigate long-term treatment effects and effects of symptom monitoring on mental health outcomes. Notably, the evidence on the use of digital interventions for the elderly and children is very limited. This is an important finding as these age groups may be particularly challenged by the current pandemic. Fourth, most studies to date do not specifically investigate the additive effects on health-related outcomes when using more advanced techniques (eg, accelerometer, GPS) to further personalize the delivery of intervention components, gamification elements, and the integration of other technologies such as wearables, although it has been described to be potentially beneficial in some of the included reviews [25,89,92,114]. Fifth, the theoretical basis of most digital interventions that have been described in previous reviews were found to be CBTs or third-wave CBTs as they may be particularly amendable to translation into digital intervention components [23,25,26]. Thus, clinicians with an expertise in CBT techniques may find it easier to purposefully incorporate intervention components delivered using digital tools in their daily clinical routines. However, findings suggest that there is a need to further improve the theoretical foundation of digital intervention, particularly mHealth interventions publicly available in major app stores. Sixth, the data available on the process quality and cost-effectiveness of eHealth and mHealth interventions are limited. Seventh, users frequently report concerns about data safety and privacy [115]. While eHealth and mHealth interventions developed and evaluated by research groups generally comply with the General Data Protection Regulation (in European countries) and work in accordance with Good Clinical Practice standards, the contents of many mHealth apps currently available in major app stores do not explicitly refer to existing clinical guidelines and recommendations by learned societies [50,116]. There are a number of reviews that have concluded that mHealth apps have problematic data-sharing and privacy practices [8,27,28] and that there may not only be a lack of quality of offered content but even harmful intervention components. In addition, although not specifically reported in included systematic reviews and meta-analyses, the recent surge in the use of popular and freely available platforms (eg, Zoom, Skype) rather than secured platforms to provide online mental health services may be another cause of concern [117] as these platforms mostly do not comply with national standards for sensitive patient data protection. In order to demonstrate user safety, clinical guidelines should be explicitly taken into account and advice by mental health professionals, learned societies, and IT (information technology) professionals actively incorporated. Overall, apps available in app stores should be used with caution due to risks in data and clinical safety as well as a lack of evidence on their effectiveness.

Limitations

This meta-review has several limitations. Because of time constraints and the rapid meta-review format of this study, the quality of included systematic reviews was not evaluated using established assessment tools (eg, the AMSTAR 2 [A Measurement Tool to Assess Systematic Reviews] checklist [118]). Along similar lines, the conclusions drawn in this meta-review on the quality of evidence are largely based on quality assessments undertaken in the included systematic reviews and meta-analyses. However, if the quality of evidence was not systematically evaluated using a standardized approach, it is indicated in Multimedia Appendix 2. Additionally, only 1 reviewer screened identified articles while a second reviewer independently screened a randomly selected subset (40%) of studies. However, this meta-review was conducted in line with the state of the art of conducting rapid reviews [119]. Furthermore, the World Health Organization has explicitly recommended rapid reviews for evidence synthesis during the ongoing public health crisis, given these are urgently needed for policy makers and the public [120].

In considering the urgent need of continued access to mental health care for vulnerable individuals during the COVID-19 pandemic, and the importance of developing and implementing public mental health prevention and promotion strategies, digital interventions should be provided by public health services and routinely offered when infection control measures are implemented during pandemics. Since there is currently no direct evidence on digital interventions that aim to minimize the psychosocial impact of previous coronavirus and influenza virus outbreaks, digital interventions should be developed and evaluated by research groups in close collaboration with relevant stakeholders to ensure established standards for investigating quality from the user perspective, effectiveness, and cost-effectiveness are met. Importantly, evidence-based digital interventions are scalable and can be rapidly delivered at the population level. This may facilitate delivering personalized care and minimizing the negative impact of the COVID-19 pandemic on public mental health.

Conclusions

Decision makers and stakeholders, including policy makers, technology companies, and public health professionals, should join forces to develop evidence-based strategies for mental health care in the area of public mental health provision, especially in moments of public health crises. As studies from previous pandemics, as well as accumulating evidence from the COVID-19 pandemic, suggest a negative impact on public mental health, the development and implementation of mental health promotion and prevention strategies at the population level may be an important measure to improve public mental

health. Digital interventions that incorporate contact with mental health staff in a blended-care approach may be particularly suited to alleviate mental health burden in help-seeking individuals. At times of COVID-19 and physical distancing measures, this may be translated into a digital blended-care approach by combining telemedical with internet-based eHealth or smartphone-based mHealth interventions. Furthermore, efforts should be made to systematically evaluate currently available digital interventions based on established criteria of digital mental health and mental health services research, as demonstrated by recent initiatives (eg, National Health Service Apps Library in the United Kingdom; Platform for Digital Health Applications in Germany; App Evaluation Database provided by the Division of Digital Psychiatry, Beth Israel Deaconess Medical Center, in the United States) [121-123]. This would systematize the search for evidence-based mHealth apps and thus allow clinicians and interested users to make more informed decisions on the quality of currently available digital interventions. There is also a need to carefully examine the role of social inequalities and the related digital divide as well as possible barriers (eg, disproportional access to necessary technologies, educational requirements, language skills, cultural factors, motor or cognitive impairments), which can influence the access to and use of the information platforms of digital mental health interventions.

Acknowledgments

The authors wish to thank research assistants Nina Müller and Viktoria Pöss for helping with data extraction. CR and UR contributed to the conception and design of the study. CR carried out the literature search and data extraction (with the help of Müller and Pöss). CR, AS, IP, and UR contributed to data interpretation and qualitative synthesis, and to the writing of the manuscript. CR and AS wrote the first draft of the manuscript. All authors contributed to critical revision of the manuscript and approved the final version. CR, AR, and UR guarantee the integrity of the work.

UR is supported by a Heisenberg professorship from the German Research Foundation (grant #389624707). The funder had no role in study design, search strategy, synthesis of findings, or writing of the review. The corresponding author had full access to all search items in the review and had final responsibility for the decision to submit for publication.

This work was carried out as part of the COVID-19 Public Health Research Network, an ad hoc consortium of more than 25 scientific societies and organizations that are active in the field of public health. They bring together their expertise in research methods, epidemiology, statistics, social sciences, demography, and medicine. The COVID-19 Public Health Research Network represents thousands of scientists from Germany, Austria, and Switzerland.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The full electronic search strategy for one database.
[DOCX File , 22 KB - jmir_v23i3e23365_app1.docx]

Multimedia Appendix 2
Complete summary of included reviews on eHealth, mHealth, and mixed interventions. Findings on target populations, intervention components, theoretical and evidence base on process/outcomes, primary outcomes and quality of evidence, secondary outcomes, quality from the user perspective, safety, and cost-effectiveness are shown.
[DOCX File , 82 KB - jmir_v23i3e23365_app2.docx]

References

1. Röhr S, Müller F, Jung F, Apfelbacher C, Seidler A, Riedel-Heller SG. [Psychosocial Impact of Quarantine Measures During Serious Coronavirus Outbreaks: A Rapid Review]. Psychiatr Prax 2020 May;47(4):179-189 [FREE Full text] [doi: 10.1055/a-1159-5562] [Medline: 32340047]
2. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. The Lancet 2020 Mar 14;395(10227):912-920. [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
3. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. The Lancet Psychiatry 2020 Jun 15;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
4. McKibbin WJ, Fernando R. The Global Macroeconomic Impacts of COVID-19: Seven Scenarios. SSRN Journal. Preprint posted online Mar 4, 2020. [doi: 10.2139/ssrn.3547729]

5. Luykx JJ, Vinkers CH, Tijdink JK. Psychiatry in Times of the Coronavirus Disease 2019 (COVID-19) Pandemic: An Imperative for Psychiatrists to Act Now. JAMA Psychiatry 2020 May 27;77(11):1097-1098. [doi: 10.1001/jamapsychiatry.2020.1225] [Medline: 32459359]
6. Yao H, Chen J, Xu Y. Patients with mental health disorders in the COVID-19 epidemic. The Lancet Psychiatry 2020 Apr;7(4):e21. [doi: 10.1016/S2215-0366(20)30090-0] [Medline: 32199510]
7. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
8. Kramer U, Borges U, Fischer F, Hoffmann W, Pobiruchin M, Vollmar HC. [DNVF-Memorandum - Health and Medical Apps]. Gesundheitswesen 2019 Oct;81(10):e154-e170. [doi: 10.1055/s-0038-1667451] [Medline: 31574556]
9. Myin-Germeys I, Klippel A, Steinhart H, Reininghaus U. Ecological momentary interventions in psychiatry. Curr Opin Psychiatry 2016 Jul;29(4):258-263. [doi: 10.1097/YCO.0000000000000255] [Medline: 27153125]
10. Rose G. Sick individuals and sick populations. Int J Epidemiol 2001 Jun;30(3):427-32; discussion 433. [doi: 10.1093/ije/30.3.427] [Medline: 11416056]
11. Wright L, Steptoe A, Fancourt D. Are we all in this together? Longitudinal assessment of cumulative adversities by socioeconomic position in the first 3 weeks of lockdown in the UK. J Epidemiol Community Health 2020 Sep;74(9):683-688 [FREE Full text] [doi: 10.1136/jech-2020-214475] [Medline: 32503892]
12. Pierce M, Hope H, Ford T, Hatch S, Hotopf M, John A, et al. Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population. The Lancet Psychiatry 2020 Oct;7(10):883-892. [doi: 10.1016/s2215-0366(20)30308-4]
13. Betsch C, Wieler L, Bosnjak M, Ramharter M, Stollorz V, Omer S. Germany COVID-19 Snapshot MOnitoring (COSMO Germany): Monitoring knowledge, risk perceptions, preventive behaviors, and public trust in the current coronavirus outbreak in Germany. PsychArchives 2020. [doi: 10.23668/psycharchives.2776]
14. Kwong A, Pearson R, Adams M, Northstone K, Tilling K, Smith D. Mental health during the COVID-19 pandemic in two longitudinal UK population cohorts. medRxiv. Preprint posted online June 18, 2020. [doi: 10.1101/2020.06.16.20133116]
15. Betsch C, Korn L, Felgendreff L, Eitze S, Schmid P, Sprengholz P, et al. German COVID-19 Snapshot MOnitoring (COSMO Germany). PsychArchives. Preprint posted online Apr 3, 2020. [FREE Full text] [doi: 10.23668/psycharchives.2863]
16. Rauschenberg C, Schick A, Goetzl C, Roehr S, Riedel-Heller S, Koppe G. Social isolation, mental health and use of digital interventions in youth during the COVID-19 pandemic: a nationally representative survey. PsyArXiv Preprints. Preprint posted online June 29, 2020. [doi: 10.31234/osf.io/v64hf]
17. Torrent-Sellens J, Díaz-Chao A, Soler-Ramos I, Saigí-Rubió F. Modelling and Predicting eHealth Usage in Europe: A Multidimensional Approach From an Online Survey of 13,000 European Union Internet Users. J Med Internet Res 2016 Jul 22;18(7):e188 [FREE Full text] [doi: 10.2196/jmir.5605] [Medline: 27450189]
18. Simons CJP, Drukker M, Evers S, van Mastrigt GAPG, Höhn P, Kramer I, et al. Economic evaluation of an experience sampling method intervention in depression compared with treatment as usual using data from a randomized controlled trial. BMC Psychiatry 2017 Dec 29;17(1):415 [FREE Full text] [doi: 10.1186/s12888-017-1577-7] [Medline: 29284448]
19. Market study on telemedicine. European Commission. 2018 Oct. URL: https://ec.europa.eu/health/sites/health/files/ehealth/docs/2018_provision_marketstudy_telemedicine_en.pdf [accessed 2021-03-01]
20. Briskman J. Sensor Tower's Q1 2020 Data Digest: Exploring COVID-19's Impact on the Global App Ecosystem. Sensor Tower Inc. 2020 Apr 15. URL: https://sensortower.com/blog/q1-2020-data-digest [accessed 2021-03-01]
21. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
22. Endnote X9 ed. Clarivate Analytics. URL: https://endnote.com/ [accessed 2021-03-01]
23. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The Benefit of Web- and Computer-Based Interventions for Stress: A Systematic Review and Meta-Analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
24. Giroux I, Goulet A, Mercier J, Jacques C, Bouchard S. Online and Mobile Interventions for Problem Gambling, Alcohol, and Drugs: A Systematic Review. Front Psychol 2017;8:954 [FREE Full text] [doi: 10.3389/fpsyg.2017.00954] [Medline: 28649211]
25. Miralles I, Granell C, Díaz-Sanahuja L, Van Woensel W, Bretón-López J, Mira A, et al. Smartphone Apps for the Treatment of Mental Disorders: Systematic Review. JMIR Mhealth Uhealth 2020 Apr 02;8(4):e14897 [FREE Full text] [doi: 10.2196/14897] [Medline: 32238332]
26. Domhardt M, Steubl L, Baumeister H. Internet- and Mobile-Based Interventions for Mental and Somatic Conditions in Children and Adolescents. Z Kinder Jugendpsychiatr Psychother 2020 Jan 13;48(1):33-46 [FREE Full text] [doi: 10.1024/1422-4917/a000625] [Medline: 30422059]
27. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: Evaluation of mental health app store quality claims. NPJ Digit Med 2019;2:18 [FREE Full text] [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]

28. Mercurio M, Larsen M, Wisniewski H, Henson P, Lagan S, Torous J. Longitudinal trends in the quality, effectiveness and attributes of highly rated smartphone health apps. Evid Based Ment Health 2020 Aug;23(3):107-111 [FREE Full text] [doi: 10.1136/ebmental-2019-300137] [Medline: 32312794]
29. Alyami M, Giri B, Alyami H, Sundram F. Social anxiety apps: a systematic review and assessment of app descriptors across mobile store platforms. Evid Based Ment Health 2017 Aug;20(3):65-70. [doi: 10.1136/eb-2017-102664] [Medline: 28666986]
30. Loucas C, Pennant M, Whittington C, Naqvi S, Sealey C, Stockton S, et al. G130 E-therapies for mental health problems in children and young people: A systematic review and focus group investigation. Archives of Disease in Childhood 2014 Apr 07;99(Suppl 1):A58-A58. [doi: 10.1136/archdischild-2014-306237.138]
31. Ennis N, Sijercic I, Monson CM. Internet-Delivered Early Interventions for Individuals Exposed to Traumatic Events: Systematic Review. J Med Internet Res 2018 Nov 14;20(11):e280 [FREE Full text] [doi: 10.2196/jmir.9795] [Medline: 30429113]
32. Elaheebocus SMRA, Weal M, Morrison L, Yardley L. Peer-Based Social Media Features in Behavior Change Interventions: Systematic Review. J Med Internet Res 2018 Feb 22;20(2):e20 [FREE Full text] [doi: 10.2196/jmir.8342] [Medline: 29472174]
33. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The Effectiveness of Prompts to Promote Engagement With Digital Interventions: A Systematic Review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
34. Lau PWC, Lau EY, Wong DP, Ransdell L. A systematic review of information and communication technology-based interventions for promoting physical activity behavior change in children and adolescents. J Med Internet Res 2011 Jul 13;13(3):e48 [FREE Full text] [doi: 10.2196/jmir.1533] [Medline: 21749967]
35. Cotie LM, Prince SA, Elliott CG, Ziss MC, McDonnell LA, Mullen KA, et al. The effectiveness of eHealth interventions on physical activity and measures of obesity among working-age women: a systematic review and meta-analysis. Obes Rev 2018 Oct;19(10):1340-1358. [doi: 10.1111/obr.12700] [Medline: 30156044]
36. Davies EB, Morriss R, Glazebrook C. Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. J Med Internet Res 2014 May 16;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
37. Flujas-Contreras JM, García-Palacios A, Gómez I. Technology-based parenting interventions for children's physical and psychological health: a systematic review and meta-analysis. Psychol Med 2019 Aug;49(11):1787-1798. [doi: 10.1017/S0033291719000692] [Medline: 30977462]
38. Harrer M, Adam SH, Baumeister H, Cuijpers P, Karyotaki E, Auerbach RP, et al. Internet interventions for mental health in university students: A systematic review and meta-analysis. Int J Methods Psychiatr Res 2019 Jun;28(2):e1759 [FREE Full text] [doi: 10.1002/mpr.1759] [Medline: 30585363]
39. Simon N, McGillivray L, Roberts NP, Barawi K, Lewis CE, Bisson JI. Acceptability of internet-based cognitive behavioural therapy (i-CBT) for post-traumatic stress disorder (PTSD): a systematic review. Eur J Psychotraumatol 2019;10(1):1646092 [FREE Full text] [doi: 10.1080/20008198.2019.1646092] [Medline: 31497259]
40. Alvarez-Jimenez M, Alcazar-Corcoles MA, González-Blanch C, Bendall S, McGorry PD, Gleeson JF. Online, social media and mobile technologies for psychosis treatment: a systematic review on novel user-led interventions. Schizophr Res 2014 Jun;156(1):96-106. [doi: 10.1016/j.schres.2014.03.021] [Medline: 24746468]
41. Rost T, Stein J, Löbner M, Kersting A, Luck-Sikorski C, Riedel-Heller SG. User Acceptance of Computerized Cognitive Behavioral Therapy for Depression: Systematic Review. J Med Internet Res 2017 Sep 13;19(9):e309 [FREE Full text] [doi: 10.2196/jmir.7662] [Medline: 28903893]
42. Biagianti B, Quraishi SH, Schlosser DA. Potential Benefits of Incorporating Peer-to-Peer Interactions Into Digital Interventions for Psychotic Disorders: A Systematic Review. Psychiatr Serv 2018 Apr 01;69(4):377-388 [FREE Full text] [doi: 10.1176/appi.ps.201700283] [Medline: 29241435]
43. Aref-Adib G, McCloud T, Ross J, O'Hanlon P, Appleton V, Rowe S, et al. Factors affecting implementation of digital health interventions for people with psychosis or bipolar disorder, and their family and friends: a systematic review. The Lancet Psychiatry 2019 Mar;6(3):257-266. [doi: 10.1016/S2215-0366(18)30302-X] [Medline: 30522979]
44. Goldberg SB, Buck B, Raphaely S, Fortney JC. Measuring Psychiatric Symptoms Remotely: a Systematic Review of Remote Measurement-Based Care. Curr Psychiatry Rep 2018 Aug 28;20(10):81. [doi: 10.1007/s11920-018-0958-z] [Medline: 30155749]
45. Faurholt-Jepsen M, Munkholm K, Frost M, Bardram JE, Kessing LV. Electronic self-monitoring of mood using IT platforms in adult patients with bipolar disorder: A systematic review of the validity and evidence. BMC Psychiatry 2016;16:7 [FREE Full text] [doi: 10.1186/s12888-016-0713-0] [Medline: 26769120]
46. Erbe D, Eichert H, Riper H, Ebert DD. Blending Face-to-Face and Internet-Based Interventions for the Treatment of Mental Disorders in Adults: Systematic Review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
47. Walsh S, Golden E, Priebe S. Systematic review of patients' participation in and experiences of technology-based monitoring of mental health symptoms in the community. BMJ Open 2016 Jun 21;6(6):e008362 [FREE Full text] [doi: 10.1136/bmjopen-2015-008362] [Medline: 27329437]

48. Pravettoni G, Triberti S. P5 eHealth: An Agenda for the Health Technologies of the Future. Cham: Springer International Publishing; 2020.
49. Hutton L, Price BA, Kelly R, McCormick C, Bandara AK, Hatzakis T, et al. Assessing the Privacy of mHealth Apps for Self-Tracking: Heuristic Evaluation Approach. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e185 [FREE Full text] [doi: 10.2196/mhealth.9217] [Medline: 30348623]
50. Arguel A, Perez-Concha O, Li SYW, Lau AYS. Theoretical approaches of online social network interventions and implications for behavioral change: a systematic review. J Eval Clin Pract 2018 Feb;24(1):212-221. [doi: 10.1111/jep.12655] [Medline: 27709724]
51. Massoudi B, Holvast F, Bockting CLH, Burger H, Blanker MH. The effectiveness and cost-effectiveness of e-health interventions for depression and anxiety in primary care: A systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:728-743. [doi: 10.1016/j.jad.2018.11.050] [Medline: 30447572]
52. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, Expert Advisory Group. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
53. Chebli J, Blaszczynski A, Gainsbury SM. Internet-Based Interventions for Addictive Behaviours: A Systematic Review. J Gambl Stud 2016 Dec;32(4):1279-1304. [doi: 10.1007/s10899-016-9599-5] [Medline: 27002522]
54. Hadjistavropoulos HD, Mehta S, Wilhelms A, Keough MT, Sundström C. A systematic review of internet-delivered cognitive behavior therapy for alcohol misuse: study characteristics, program content and outcomes. Cogn Behav Ther 2020 Jul;49(4):327-346. [doi: 10.1080/16506073.2019.1663258] [Medline: 31599198]
55. O'Rourke L, Humphris G, Baldacchino A. Electronic communication based interventions for hazardous young drinkers: A systematic review. Neurosci Biobehav Rev 2016 Dec;68:880-890. [doi: 10.1016/j.neubiorev.2016.07.021] [Medline: 27453570]
56. Tait RJ, Christensen H. Internet-based interventions for young people with problematic substance use: a systematic review. Med J Aust 2010 Jun 07;192(S11):S15-S21. [doi: 10.5694/j.1326-5377.2010.tb03687.x] [Medline: 20528701]
57. Boumparis N, Loheide-Niesmann L, Blankers M, Ebert DD, Korf D, Schaub MP, et al. Short- and long-term effects of digital prevention and treatment interventions for cannabis use reduction: A systematic review and meta-analysis. Drug Alcohol Depend 2019 Jul 01;200:82-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2019.03.016] [Medline: 31112834]
58. Asuzu K, Matin A, Van Noord M, Onigu-Otite E. Electronically-Delivered Interventions to Reduce Cannabis Use in Adolescents: A Systematic Review. APS 2019 Jan 31;8(3):195-213. [doi: 10.2174/2210676609666181203094249]
59. Deady M, Choi I, Calvo RA, Glozier N, Christensen H, Harvey SB. eHealth interventions for the prevention of depression and anxiety in the general population: a systematic review and meta-analysis. BMC Psychiatry 2017 Aug 29;17(1):310 [FREE Full text] [doi: 10.1186/s12888-017-1473-1] [Medline: 28851342]
60. Richardson T, Stallard P, Velleman S. Computerised cognitive behavioural therapy for the prevention and treatment of depression and anxiety in children and adolescents: a systematic review. Clin Child Fam Psychol Rev 2010 Sep;13(3):275-290. [doi: 10.1007/s10567-010-0069-9] [Medline: 20532980]
61. Fleming TM, Cheek C, Merry SN, Thabrew H, Bridgman H, Stasiak K, et al. Juegos serios para el tratamiento o la prevención de la depresión: una revisión sistemática. RPPC 2014;19(3):227-242. [doi: 10.5944/rppc.vol.19.num.3.2014.13904]
62. Lau HM, Smit JH, Fleming TM, Riper H. Serious Games for Mental Health: Are They Accessible, Feasible, and Effective? A Systematic Review and Meta-analysis. Front Psychiatry 2016;7:209 [FREE Full text] [doi: 10.3389/fpsyt.2016.00209] [Medline: 28149281]
63. Grist R, Croker A, Denne M, Stallard P. Technology Delivered Interventions for Depression and Anxiety in Children and Adolescents: A Systematic Review and Meta-analysis. Clin Child Fam Psychol Rev 2019 Jun;22(2):147-171 [FREE Full text] [doi: 10.1007/s10567-018-0271-8] [Medline: 30229343]
64. Baker S, Warburton J, Waycott J, Batchelor F, Hoang T, Dow B, et al. Combatting social isolation and increasing social participation of older adults through the use of technology: A systematic review of existing evidence. Australas J Ageing 2018 Sep;37(3):184-193. [doi: 10.1111/ajag.12572] [Medline: 30022583]
65. Grist R, Cavanagh K. Computerised Cognitive Behavioural Therapy for Common Mental Health Disorders, What Works, for Whom Under What Circumstances? A Systematic Review and Meta-analysis. J Contemp Psychother 2013 Sep 4;43(4):243-251. [doi: 10.1007/s10879-013-9243-y]
66. Păsărelu CR, Andersson G, Bergman Nordgren L, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017 Jan;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
67. Ye X, Bapuji SB, Winters SE, Struthers A, Raynard M, Metge C, et al. Effectiveness of internet-based interventions for children, youth, and young adults with anxiety and/or depression: a systematic review and meta-analysis. BMC Health Serv Res 2014 Jul 18;14:313 [FREE Full text] [doi: 10.1186/1472-6963-14-313] [Medline: 25037951]
68. Stech EP, Lim J, Upton EL, Newby JM. Internet-delivered cognitive behavioral therapy for panic disorder with or without agoraphobia: a systematic review and meta-analysis. Cogn Behav Ther 2020 Jul;49(4):270-293. [doi: 10.1080/16506073.2019.1628808] [Medline: 31303121]

69. Rice SM, Goodall J, Hetrick SE, Parker AG, Gilbertson T, Amminger GP, et al. Online and social networking interventions for the treatment of depression in young people: a systematic review. J Med Internet Res 2014 Sep 16;16(9):e206 [FREE Full text] [doi: 10.2196/jmir.3304] [Medline: 25226790]
70. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
71. Wright JH, Owen JJ, Richards D, Eells TD, Richardson T, Brown GK, et al. Computer-Assisted Cognitive-Behavior Therapy for Depression: A Systematic Review and Meta-Analysis. J Clin Psychiatry 2019 Mar 19;80(2) [FREE Full text] [doi: 10.4088/JCP.18r12188] [Medline: 30900849]
72. Sikorski C, Luppa M, Kersting A, König HH, Riedel-Heller SG. [Computer-aided cognitive behavioral therapy for depression]. Psychiatr Prax 2011 Feb;38(2):61-68. [doi: 10.1055/s-0030-1248575] [Medline: 20972949]
73. Xiang X, Wu S, Zuverink A, Tomasino KN, An R, Himle JA. Internet-delivered cognitive behavioral therapies for late-life depressive symptoms: a systematic review and meta-analysis. Aging Ment Health 2020 Aug 27;24(8):1196-1206. [doi: 10.1080/13607863.2019.1590309] [Medline: 30913898]
74. Danielsson A, Eriksson A, Allebeck P. Technology-based support via telephone or web: a systematic review of the effects on smoking, alcohol use and gambling. Addict Behav 2014 Dec;39(12):1846-1868. [doi: 10.1016/j.addbeh.2014.06.007] [Medline: 25128637]
75. Lin LA, Casteel D, Shigekawa E, Weyrich MS, Roby DH, McMenamin SB. Telemedicine-delivered treatment interventions for substance use disorders: A systematic review. J Subst Abuse Treat 2019 Jun;101:38-49. [doi: 10.1016/j.jsat.2019.03.007] [Medline: 31006553]
76. Gilmore AK, Wilson SM, Skopp NA, Osenbach JE, Reger G. A systematic review of technology-based interventions for co-occurring substance use and trauma symptoms. J Telemed Telecare 2017 Sep;23(8):701-709. [doi: 10.1177/1357633X16664205] [Medline: 27534823]
77. Pittock A, Hodges L, Lawrie SM. The effectiveness of internet-delivered cognitive behavioural therapy for those with bulimic symptoms: a systematic review : A review of iCBT treatment for bulimic symptoms. BMC Res Notes 2018 Oct 22;11(1):748 [FREE Full text] [doi: 10.1186/s13104-018-3843-2] [Medline: 30348226]
78. Meyer TD, Casarez R, Mohite SS, La Rosa N, Iyengar MS. Novel technology as platform for interventions for caregivers and individuals with severe mental health illnesses: A systematic review. J Affect Disord 2018 Jan 15;226:169-177. [doi: 10.1016/j.jad.2017.09.012] [Medline: 28987999]
79. Sierra M, Ruiz F, Flórez C. A Systematic Review and Meta-Analysis of Third-Wave Online Interventions for Depression. RLP 2018 Sep 30;50(2):1641-1648. [doi: 10.14349/rlp.2018.v50.n2.6]
80. Berryhill MB, Halli-Tierney A, Culmer N, Williams N, Betancourt A, King M, et al. Videoconferencing psychological therapy and anxiety: a systematic review. Fam Pract 2019 Jan 25;36(1):53-63. [doi: 10.1093/fampra/cmy072] [Medline: 30188992]
81. Castro A, Gili M, Ricci-Cabello I, Roca M, Gilbody S, Perez-Ara MA, et al. Effectiveness and adherence of telephone-administered psychotherapy for depression: A systematic review and meta-analysis. J Affect Disord 2020 Jan 01;260:514-526. [doi: 10.1016/j.jad.2019.09.023] [Medline: 31539688]
82. Coughtrey AE, Pistrang N. The effectiveness of telephone-delivered psychological therapies for depression and anxiety: A systematic review. J Telemed Telecare 2016 Dec 30;24(2):65-74. [doi: 10.1177/1357633x16686547]
83. Rees C, Maclaine E. A Systematic Review of Videoconference-Delivered Psychological Treatment for Anxiety Disorders. Australian Psychologist 2015 Jul 20;50(4):259-264. [doi: 10.1111/ap.12122]
84. Bolton AJ, Dorstyn DS. Telepsychology for Posttraumatic Stress Disorder: A systematic review. J Telemed Telecare 2015 Jul;21(5):254-267. [doi: 10.1177/1357633X15571996] [Medline: 25712113]
85. Lewis C, Roberts NP, Simon N, Bethell A, Bisson JI. Internet-delivered cognitive behavioural therapy for post-traumatic stress disorder: systematic review and meta-analysis. Acta Psychiatr Scand 2019 Dec;140(6):508-521. [doi: 10.1111/acps.13079] [Medline: 31359407]
86. Baker AL, Turner A, Beck A, Berry K, Haddock G, Kelly PJ, et al. Telephone-delivered psychosocial interventions targeting key health priorities in adults with a psychotic disorder: systematic review. Psychol Med 2018 Dec 25;48(16):2637-2657 [FREE Full text] [doi: 10.1017/S0033291718001125] [Medline: 29798730]
87. Gentry MT, Lapid MI, Clark MM, Rummans TA. Evidence for telehealth group-based treatment: A systematic review. J Telemed Telecare 2019 Jul;25(6):327-342. [doi: 10.1177/1357633X18775855] [Medline: 29788807]
88. Irvine A, Drew P, Bower P, Brooks H, Gellatly J, Armitage CJ, et al. Are there interactional differences between telephone and face-to-face psychological therapy? A systematic review of comparative studies. J Affect Disord 2020 Mar 15;265:120-131 [FREE Full text] [doi: 10.1016/j.jad.2020.01.057] [Medline: 32090733]
89. Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? Ten strategies for increasing the timeliness and usefulness of eHealth research. J Med Internet Res 2014 Feb 19;16(2):e36 [FREE Full text] [doi: 10.2196/jmir.2925] [Medline: 24554442]
90. Böhm B, Karwiese SD, Böhm H, Oberhoffer R. Effects of Mobile Health Including Wearable Activity Trackers to Increase Physical Activity Outcomes Among Healthy Children and Adolescents: Systematic Review. JMIR Mhealth Uhealth 2019 Apr 30;7(4):e8298 [FREE Full text] [doi: 10.2196/mhealth.8298] [Medline: 31038460]

91. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
92. Feter N, Dos Santos TS, Caputo EL, da Silva MC. What is the role of smartphones on physical activity promotion? A systematic review and meta-analysis. Int J Public Health 2019 Jun;64(5):679-690. [doi: 10.1007/s00038-019-01210-7] [Medline: 30758514]
93. Kim H, Seo K. Smartphone-Based Health Program for Improving Physical Activity and Tackling Obesity for Young Adults: A Systematic Review and Meta-Analysis. Int J Environ Res Public Health 2019 Dec 18;17(1):15 [FREE Full text] [doi: 10.3390/ijerph17010015] [Medline: 31861359]
94. Rose T, Barker M, Maria Jacob C, Morrison L, Lawrence W, Strömmer S, et al. A Systematic Review of Digital Interventions for Improving the Diet and Physical Activity Behaviors of Adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
95. Muntaner A, Vidal-Conti J, Palou P. Increasing physical activity through mobile device interventions: A systematic review. Health Informatics J 2016 Sep 3;22(3):451-469 [FREE Full text] [doi: 10.1177/1460458214567004] [Medline: 25649783]
96. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
97. Loo Gee B, Griffiths KM, Gulliver A. Effectiveness of mobile technologies delivering Ecological Momentary Interventions for stress and anxiety: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):221-229 [FREE Full text] [doi: 10.1093/jamia/ocv043] [Medline: 25997643]
98. Holmes NA, van Agteren JE, Dorstyn DS. A systematic review of technology-assisted interventions for co-morbid depression and substance use. J Telemed Telecare 2019 Apr;25(3):131-141. [doi: 10.1177/1357633X17748991] [Medline: 29310532]
99. Rathbone AL, Prescott J. The Use of Mobile Apps and SMS Messaging as Physical and Mental Health Interventions: Systematic Review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
100. Rootes-Murdy K, Glazer KL, Van Wert MJ, Mondimore FM, Zandi PP. Mobile technology for medication adherence in people with mood disorders: A systematic review. J Affect Disord 2018 Feb;227:613-617. [doi: 10.1016/j.jad.2017.11.022] [Medline: 29172054]
101. Dick S, Whelan E, Davoren MP, Dockray S, Heavin C, Linehan C, et al. A systematic review of the effectiveness of digital interventions for illicit substance misuse harm reduction in third-level students. BMC Public Health 2019 Sep 09;19(1):1244 [FREE Full text] [doi: 10.1186/s12889-019-7583-6] [Medline: 31500618]
102. Song T, Qian S, Yu P. Mobile Health Interventions for Self-Control of Unhealthy Alcohol Use: Systematic Review. JMIR Mhealth Uhealth 2019 Jan 29;7(1):e10899 [FREE Full text] [doi: 10.2196/10899] [Medline: 30694200]
103. Yonker LM, Zan S, Scirica CV, Jethwani K, Kinane TB. "Friending" teens: systematic review of social media in adolescent and young adult health care. J Med Internet Res 2015 Jan 05;17(1):e4 [FREE Full text] [doi: 10.2196/jmir.3692] [Medline: 25560751]
104. Arshad U, Farhat-Ul-Ain, Gauntlett J, Husain N, Chaudhry N, Taylor PJ. A Systematic Review of the Evidence Supporting Mobile- and Internet-Based Psychological Interventions For Self-Harm. Suicide Life Threat Behav 2020 Feb;50(1):151-179 [FREE Full text] [doi: 10.1111/sltb.12583] [Medline: 31448847]
105. Clarke S, Hanna D, Mulholland C, Shannon C, Urquhart C. A systematic review and meta-analysis of digital health technologies effects on psychotic symptoms in adults with psychosis. Psychosis 2019 Aug 12;11(4):362-373. [doi: 10.1080/17522439.2019.1632376]
106. Kreuze E, Jenkins C, Gregoski M, York J, Mueller M, Lamis DA, et al. Technology-enhanced suicide prevention interventions: A systematic review. J Telemed Telecare 2017 Jul;23(6):605-617. [doi: 10.1177/1357633X16657928] [Medline: 27377792]
107. Perry Y, Werner-Seidler A, Calear AL, Christensen H. Web-Based and Mobile Suicide Prevention Interventions for Young People: A Systematic Review. J Can Acad Child Adolesc Psychiatry 2016;25(2):73-79 [FREE Full text] [Medline: 27274742]
108. Ilagan GS, Iliakis EA, Wilks CR, Vahia IV, Choi-Kain LW. Smartphone applications targeting borderline personality disorder symptoms: a systematic review and meta-analysis. Borderline Personal Disord Emot Dysregul 2020;7:12 [FREE Full text] [doi: 10.1186/s40479-020-00127-5] [Medline: 32549987]
109. Sander LB, Schorndanner J, Terhorst Y, Spanhel K, Pryss R, Baumeister H, et al. 'Help for trauma from the app stores?' A systematic review and standardised rating of apps for Post-Traumatic Stress Disorder (PTSD). Eur J Psychotraumatol 2020;11(1):1701788 [FREE Full text] [doi: 10.1080/20008198.2019.1701788] [Medline: 32002136]
110. Sucala M, Cuijpers P, Muench F, Cardo R, Soflau R, Dobrean A, et al. Anxiety: There is an app for that. A systematic review of anxiety apps. Depress Anxiety 2017 Jun;34(6):518-525. [doi: 10.1002/da.22654] [Medline: 28504859]
111. Terhorst Y, Rathner E, Baumeister H, Sander L. «Hilfe aus dem App-Store?»: Eine systematische Übersichtsarbeit und Evaluation von Apps zur Anwendung bei Depressionen. Verhaltenstherapie 2018 May 8;28(2):101-112. [doi: 10.1159/000481692]

112. Axelsson E, Hedman-Lagerlöf E. Cognitive behavior therapy for health anxiety: systematic review and meta-analysis of clinical efficacy and health economic outcomes. Expert Rev Pharmacoecon Outcomes Res 2019 Dec;19(6):663-676. [doi: 10.1080/14737167.2019.1703182] [Medline: 31859542]
113. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
114. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
115. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of Interventions Delivered Online and Through Mobile Phones for People Who Experience Severe Mental Health Problems: A Systematic Review. J Med Internet Res 2016 May 31;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]
116. Mantovani E, Antokol J, Hoekstra M, Nouwt S, Schutte N, Zilgalvis P. Towards a Code of Conduct on Privacy for mHealth to Foster Trust Amongst Users of Mobile Health Applications. In: Leenes R, van Brakel R, Gutwirth S, De Hert P, editors. Data Protection and Privacy: (In)visibilities and Infrastructures. Cham: Springer International Publishing; 2017:81-106.
117. Soron TR, Shariful Islam SM, Ahmed HU, Ahmed SI. The hope and hype of telepsychiatry during the COVID-19 pandemic. The Lancet Psychiatry 2020 Aug;7(8):e50 [FREE Full text] [doi: 10.1016/S2215-0366(20)30260-1] [Medline: 32711713]
118. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
119. Tricco AC, Antony J, Zarin W, Strifler L, Ghassemi M, Ivory J, et al. A scoping review of rapid review methods. BMC Med 2015 Sep 16;13:224 [FREE Full text] [doi: 10.1186/s12916-015-0465-6] [Medline: 26377409]
120. Rapid reviews to strengthen health policy and systems: a practical guide. World Health Organization, Alliance for Health Policy and Systems Research. 2017. URL: https://www.who.int/alliance-hpsr/resources/publications/rapid-review-guide/en/ [accessed 2021-03-01]
121. Digital health applications (DiGA). Federal Institute for Drugs and Medical Devices (BfArM), Germany. 2020. URL: https://www.bfarm.de/EN/MedicalDevices/DiGA/_node.html [accessed 2021-03-01]
122. Apps Library - Category Mental Health. United Kingdom National Health Service. 2020. URL: https://www.nhs.uk/apps-library/filter/?categories=Mental%20health [accessed 2021-03-01]
123. App Evaluation Database. Division of Digital Psychiatry, Beth Israel Deaconess Medical Center. 2020. URL: https://apps.digitalpsych.org [accessed 2021-03-01]

Abbreviations

AMSTAR 2: A Measurement Tool to Assess Systematic Reviews
CBT: cognitive-behavioral therapy
COSMO Germany: German COVID-19 Snapshot Monitoring
IT: information technology
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach, C Basch; submitted 10.08.20; peer-reviewed by B Buck, A Beck, A Salzwedel, S Elaheebocus; comments to author 19.09.20; revised version received 14.02.21; accepted 17.02.21; published 10.03.21.

Please cite as:
Rauschenberg C, Schick A, Hirjak D, Seidler A, Paetzold I, Apfelbacher C, Riedel-Heller SG, Reininghaus U
Evidence Synthesis of Digital Interventions to Mitigate the Negative Impact of the COVID-19 Pandemic on Public Mental Health: Rapid Meta-review
J Med Internet Res 2021;23(3):e23365
URL: https://www.jmir.org/2021/3/e23365
doi:10.2196/23365
PMID:33606657

Original Paper

Factors Affecting Public Adoption of COVID-19 Prevention and Treatment Information During an Infodemic: Cross-sectional Survey Study

Yangyang Han[1*], BS; Binshan Jiang[1*], BS; Rui Guo[1], PhD

School of Public Health, Capital Medical University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Rui Guo, PhD
School of Public Health
Capital Medical University
No.10 Youanmenwai, Xitoutiao
Beijing
China
Phone: 86 01083911573
Email: guorui@ccmu.edu.cn

Abstract

Background: With the spread of COVID-19, an infodemic is also emerging. In public health emergencies, the use of information to enable disease prevention and treatment is incredibly important. Although both the information adoption model (IAM) and health belief model (HBM) have their own merits, they only focus on information or public influence factors, respectively, to explain the public's intention to adopt online prevention and treatment information.

Objective: The aim of this study was to fill this gap by using a combination of the IAM and the HBM as the framework for exploring the influencing factors and paths in public health events that affect the public's adoption of online health information and health behaviors, focusing on both objective and subjective factors.

Methods: We carried out an online survey to collect responses from participants in China (N=501). Structural equation modeling was used to evaluate items, and confirmatory factor analysis was used to calculate construct reliability and validity. The goodness of fit of the model and mediation effects were analyzed.

Results: The overall fitness indices for the model developed in this study indicated an acceptable fit. Adoption intention was predicted by information characteristics (β=.266, *P*<.001) and perceived usefulness (β=.565, *P*<.001), which jointly explained nearly 67% of the adoption intention variance. Information characteristics (β=.244, *P*<.001), perceived drawbacks (β=–.097, *P*=.002), perceived benefits (β=.512, *P*<.001), and self-efficacy (β=.141, *P*<.001) jointly determined perceived usefulness and explained about 81% of the variance of perceived usefulness. However, social influence did not have a statistically significant impact on perceived usefulness, and self-efficacy did not significantly influence adoption intention directly.

Conclusions: By integrating IAM and HBM, this study provided the insight and understanding that perceived usefulness and adoption intention of online health information could be influenced by information characteristics, people's perceptions of information drawbacks and benefits, and self-efficacy. Moreover, people also exhibited proactive behavior rather than reactive behavior to adopt information. Thus, we should consider these factors when helping the *informed public* obtain useful information via two approaches: one is to improve the quality of government-based and other official information, and the other is to improve the public's capacity to obtain information, in order to promote truth and fight rumors. This will, in turn, contribute to saving lives as the pandemic continues to unfold and run its course.

(J Med Internet Res 2021;23(3):e23097) doi:10.2196/23097

KEYWORDS

information adoption; infodemic; China; health information; infodemiology; COVID-19; public health

Introduction

Background

The global outbreak and rapid spread of the COVID-19 pandemic has led the world's public health and safety systems to face great challenges. As of April 16, 2020, the COVID-19 pandemic has affected at least 211 countries, with 2,362,704 confirmed cases and more than 165,006 deaths globally, while China has confirmed 82,367 cases and the cumulative number of deaths was 3342 [1]. The disease rapidly spread in China because of the massive population migration (ie, Chunyun, also known as the Spring Festival travel rush, which refers to a human migration during this festival) and lack of prevention and control information during the early stage of the pandemic. Moreover, the World Health Organization (WHO) has not only signaled the health risks of COVID-19 but has also labeled the situation as an infodemic, due to the amount of information, both truth and rumors, circulating around this topic [2,3].

Although public health information dissemination represents an exciting combination of broadcasting, sharing, and retrieving relevant health information [4,5], the COVID-19 infodemic did not come as a surprise [6]. The massive growth of health information on the internet was seen to be a real problem [4]. Too much information makes it difficult to find trustworthy sources of information and may harm people's health. Therefore, the quality of online health information is essential, especially when truth and rumors were still intertwined, which caused the infodemic. However, in the process of fighting COVID-19, the WHO and health authorities worldwide have been working closely with social media platforms, including Facebook, Google, Twitter, and YouTube, to provide evidence-based information to the general public in order to actively counter the rumors that have been circulating [7]. Prevention and treatment information about COVID-19 continues to spread and has an influence on populations.

At the same time, the lack of transparent, timely, and effective risk communication by health authorities regarding this emerging infectious disease in its early stage failed to bring about the appropriate level of public awareness and behavioral responses, such as avoidance of mass gatherings and personal protection in China, Europe, and the United States [8]. Online guidelines providing information for prevention and treatment are inconsistent in adapting to new knowledge, and changing or conflicting information can also confuse the public [9].

Since the beginning of the COVID-19 pandemic, information consumption has increased rapidly and significantly [10]. The new generation of health care consumers consists of an *informedpublic*, gleaning truth and rumors about health information with both positive and negative effects on themselves [11]. During periods of SARS spread, most people obtained SARS information from television [12]. By the time the Zika virus emerged in 2015-2016, Google Trends showed a significant increase in public searches on the internet related to the Zika virus [13], and the number of searches on video platforms, such as YouTube, had also increased rapidly [14]. Internet-based platforms that people utilized became diversified. During the COVID-19 pandemic, social distancing and stay-at-home restrictions caused the public to be fully exposed to social media; during this time, people actively searched or passively received a large amount of health-related information to prevent and treat diseases. In China, social media platforms, such as WeChat and Douyin (ie, TikTok), played an essential role in obtaining information after the virus began spreading [15]. However, studies have confirmed that over one-quarter of the most viewed YouTube videos about COVID-19 contained rumors, and these reached millions of viewers worldwide [16]. Therefore, determining whether people can use the prevention and treatment information they find on social media is critical in this pandemic.

Moreover, identifying health information from prevention and control measures is a major blind spot for the public. The public are partly responsible for selecting and filtering trustworthy health information [17]. Research shows that more than half of respondents trust almost all information online [18], and people are more likely to believe health rumors because of basic safety needs [19]. As the core part of health behavior theory, behavior intention is the subjective possibility of engaging in certain behavior. Some studies have confirmed that variables in the health belief model (HBM) can materially affect information adoption intention, and this process will affect subsequent health behaviors [20].

Therefore, if concern about identifying trustworthy information is reflected in the global population, we believe that the influencing factors in the subjective and objective aspects of engaging in that behavior may affect information adoption. Motivated by previous studies, this study was based on the information adoption model (IAM) and the HBM. The aim of this study was to explore the influencing factors and paths during public health events that affect the public's intention to adopt online prevention and treatment information under the infodemic. We aim to provide a basis for decision making and policy suggestions in order to deal with online health information governance in the internet era. Moreover, this study adds to the sparse literature on information adoption.

Research Model and Hypotheses

Information Adoption Model

Sussman and Siegal first proposed the IAM based on the technology acceptance model (TAM) and the theoretical perspective of the elaboration likelihood model (ELM), which regards how information influences people's decision making as the process of information adoption [21]. From the TAM, a critical aspect of how individuals act on an advocated issue or behavior is the extent to which they believe the information contained within a message is useful [21,22]. From the ELM, this process depends on elaboration likelihood, and two likely antecedents of usefulness have been suggested from this stream of research as well as two key internal validity factors [21]. The ELM explains how individuals adopt information and then change their will and behavior. Moreover, recent literature has also demonstrated that this model can be applied in the context of online information acceptance, argument quality, and source credibility; these are taken as the direct objects of information adoption, and their influence has been repeatedly verified [23,24]. Simultaneously, individuals' perceived information

usefulness based on information quality and source characteristics plays a crucial intermediary role in information adoption. Health information as a type of information fits into the IAM's influencing factors. The essence of information adoption is when individuals are persuaded by the received information and then accept the opinions or propositions expressed in the information.

Health Belief Model

The HBM has been widely used to explain preventive health behavior [25] and is one of the first and the best-known social cognition models. It focuses on the relationship of health behaviors, practices, and utilization of health services. From its initial design to predict behavioral response to the treatment received by chronic patients [26], it has been validated in different studies. Contemporary research studies have recently focused on the general health behavior of the population [27,28]. The core concept of HBM is people's perception of disease threat and an assessment of their behavior. The assessment of behavior includes evaluating the effectiveness of behavior, the input and outcome of behavior change, and the obstacle to its implementation [28].

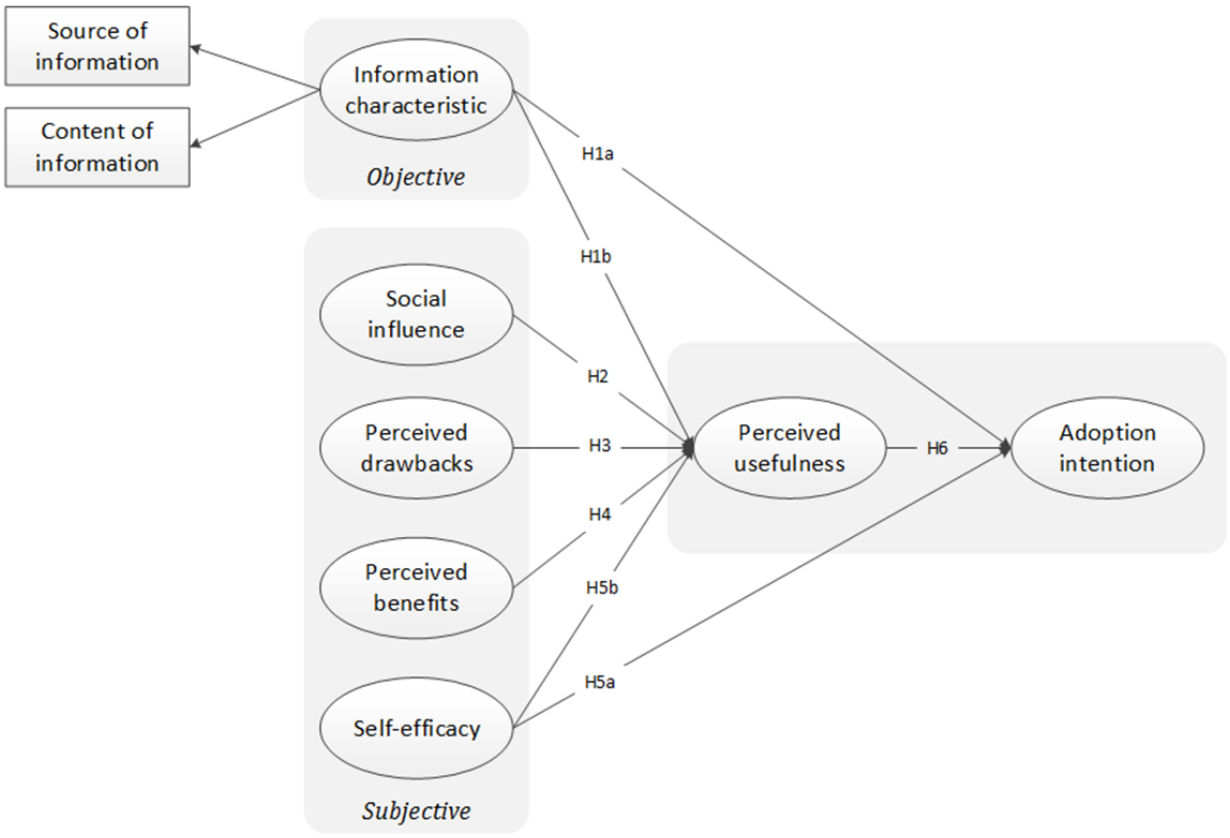
Furthermore, researchers added *cues to action*, meaning the stimuli or cues that catalyze action. Cues are divided into external cues, such as mass media, and internal cues, such as physical discomfort, that limit people's beliefs about behavioral health consequences and behavioral effects. The HBM has been widely used in health behavior change [20], which provides scientific theoretical support for understanding the impact of health information propagation on the audience's health behavior. It has become one of the most comprehensive models to understand health-related behaviors and why people undertake, or do not undertake, actions to prevent or control illnesses [29].

Integrated Model of IAM and HBM

Although IAM and HBM are commonly used models, the use of these models independently has not fully explained online health information adoption behavior. IAM focuses on the influence of information characteristics on information adoption without considering the individual's subjective status quo. However, the information's influence might change from person to person; the same content can evoke differing notions among receivers [30]. Also, HBM only considers the public's cognitive information to predict general health behavior without influencing the process. More specifically, this study argues that an individual's motivation to adopt health information will depend on the individual's subjective and information-related objective factors. It is also supported by Erkan and Evans' information acceptance model (IACM), in which a conceptual model was developed based on the integration of IAM and the theory of reasoned action; this model confirmed the influences of both information adoption and attitude toward information on consumers' purchase intentions and the influence of information usefulness on information adoption [31]. In IACM, information quality, information credibility, and information needs were all found to affect information usefulness [31]. In addition, Ahadzadeh et al combined the TAM and HBM to study health-related internet use [32]. This study demonstrated that individuals who perceived their health to be at risk, or were motivated to use the internet when they believed that the internet was useful for providing health and health management information, would be expected to have a positive attitude toward internet use for health purposes [32].

Our research mainly focused on information adoption behavior during the COVID-19 infodemic; therefore, we developed the IAM-HBM model by considering information characteristics (ie, objective factors) and the public's health beliefs (ie, subjective factors) about information. Also, the HBM has been criticized for not considering environmental factors, such as social influence (ie, friends, family, and individuals' internet providers), that can influence health-related behavior [33]. Therefore, we proposed the following path: social influence impacts perceived usefulness of information. Figure 1 shows the conceptual model proposed by this study, which incorporates information characteristics, social influence, perceived drawbacks, perceived benefits, self-efficacy, perceived usefulness, and adoption intention.

Figure 1. Integrated model based on the information adoption model and health belief model. H: hypothesis.



Hypotheses

Information characteristics refer to whether the prevention and treatment information about COVID-19 to which individuals are exposed from the internet is persuasive and to how individuals perceive the credibility of the information source, including the content and source attributes of the information. Sussman and Siegal believed that information characteristics, including quality and source, impacted people's perceived information usefulness, which then affected information adoption; they also believed that information characteristics impacted individuals' information adoption directly, which was measured by information adoption intention [21]. Thus, information characteristics are applied to our IAM-HBM model as verified factors in the IAM, impacting individuals' perceived usefulness of COVID-19 prevention and treatment information. Based on this, we hypothesized the following:

1. Hypothesis 1a. Information characteristics are positively associated with the intention to adopt COVID-19 prevention and treatment information.
2. Hypothesis 1b. Information characteristics are positively associated with perceived usefulness of COVID-19 prevention and treatment information.

Social influence refers to how individuals perceive the influence of those around them when they adopt COVID-19 prevention and control information. Venkatesh et al defined social influence as the degree to which users were affected by people around them using new technologies and systems; this included integrated subjective norms, social factors, and images [34]. The construct of social influence was added to the extended HBM to enhance the prediction ability of the model. Perceived usefulness was mediated by external variables, including social influence [35,36].

Perceived drawbacks refer to the difficulties that individuals may encounter when adopting the prevention and treatment information of COVID-19 on the internet. Such difficulties become apparent by the predicted cost of adopting healthy behaviors, including tangible and intangible costs. Tangible cost refers to the cost of perceived usefulness of information, usually measured in monetary terms. Intangible cost refers to the effort required to confirm the usefulness of information, such as time and energy [37,38]. Based on HBM, perceived drawbacks were confirmed as the most powerful single predictor of intended expectations [39]. Yun built an integrated model and verified that perceived drawbacks affect people's actions in seeking health information through perceived usefulness [40]. If social media users do not need to spend too much money or physical and mental energy, they can reduce the cost of using technology, and their total perceived usefulness will increase when searching for online health information [41].

Perceived benefits refer to the benefits that individuals may give themselves if they adopt the prevention and treatment information about COVID-19 on the internet. In the research on the HBM, Rosenstock proposed that individuals would weigh the effectiveness of behaviors through cost-benefit analysis when they adopt healthy behaviors. The perception of benefits provided a more preferential action path [41,42]. The positive experiences gained by individuals from acquiring and adopting health information behavior will promote the overall value

perception of this behavior. This will further strengthen the intention of continuing to search for health information [41]. Building on findings by Rosenstock, we explored whether perceived benefits are associated with perceived usefulness. Thus, perceived benefits were applied to this IAM-HBM model as a factor that estimates individuals' beliefs about the usefulness of COVID-19 prevention and treatment information.

Self-efficacy refers to the level of confidence in an individual's ability to prevent and treat COVID-19. As an essential part of social cognitive theory, self-efficacy refers to people's confidence in performing a specific behavior. The higher the expectations, the higher the tendency to make more considerable efforts. This concept has been widely used to understand health behavior change. Self-efficacy also had an impact on health information–seeking behavior [43]. Information on the internet regarding people's health interventions significantly improved the specific behaviors of self-efficacy, including physical exercise and healthy eating. Physiological and social advantages caused people to have more positive behavior change expectations.

Yun's integrated model demonstrated that internet self-efficacy affected users' actions through perceived usefulness when seeking health information on the internet [40]. We gravitate toward the idea that self-efficacy is a determinant of adoption intention.

Based on the discussion above, we hypothesized the following:

1. Hypothesis 2. Social influence is positively associated with perceived usefulness of COVID-19 prevention and treatment information.
2. Hypothesis 3. Perceived drawbacks are negatively associated with perceived usefulness of COVID-19 prevention and treatment information.
3. Hypothesis 4. Perceived benefits are positively associated with perceived usefulness of COVID-19 prevention and treatment information.
4. Hypothesis 5a. Self-efficacy is positively associated with COVID-19 prevention and treatment information adoption.
5. Hypothesis 5b. Self-efficacy is positively associated with perceived usefulness of COVID-19 prevention and treatment information.

Perceived usefulness refers to an individual's blanket perception of COVID-19 prevention and treatment information on the internet. It is plausible that adopting such healthy behaviors can meet individuals' health needs and help them achieve healthy outcomes. Usefulness, utility, and perceived usefulness, which were first applied to TAM by Davis et al, are used to evaluate the utility of information-seeking behavior [22]. As the crucial variable of IAM, perceived usefulness has a significant influence on adoption intention. Thus, perceived usefulness is applied to this IAM-HBM model as a factor that could influence individuals' COVID-19 prevention and treatment information adoption. In our paper, Hypothesis 6 states that perceived usefulness is positively associated with the intention to adopt COVID-19 prevention and treatment information.

Methods

Data Collection and Participants

Online questionnaires were powered by the survey platform WJX (Changsha Ranxing Information Technology Co), whose web application was embedded into social media platforms from March 24 to April 5, 2020. The electronic version of the questionnaire was uploaded to the WJX web application; respondents (ie, Chinese people in China) could fill in, submit, and share the questionnaire using a Quick Response (QR) code or using a forwarding link issued by the WJX web application. The data were collected using snowball sampling through repetitive one-to-many sharing on social media, a nonprobability sampling method; there were 528 respondents from 30 provinces in China. We gathered data using an online survey because of public space restrictions and because netizens were potentially exposed to the infodemic. After eliminating invalid responses through data filtering, 501 valid questionnaires out of 528 remained (94.9% validity rate). Table 1 shows the demographic characteristics of the sample population.

Table 1. Demographic characteristics of participants.

Characteristic	Value (N=501), n (%)
Gender	
Male	216 (43.1)
Female	285 (56.9)
Age (years)	
18-19	6 (1.2)
20-29	154 (30.7)
30-39	122 (24.4)
40-49	81 (16.2)
50-59	77 (15.4)
60-69	51 (10.2)
70-79	8 (1.5)
≥80	2 (0.4)
Education	
Junior high school diploma or below	21 (4.2)
Senior high school diploma	66 (13.2)
College graduate	84 (16.8)
Bachelor's degree	256 (51.1)
Master's degree or above	74 (14.8)
Occupation	
Student	104 (20.8)
Officer[a]	50 (10.0)
Enterprise manager	67 (13.4)
Office staff or clerk	62 (12.4)
Professional[b]	75 (15.0)
Worker or laborer	24 (4.8)
Business service	14 (2.8)
Self-employed	12 (2.4)
Freelancer	22 (4.4)
Farmer	2 (0.4)
Retired	63 (12.6)
No profession[c]	6 (1.2)

[a]Officer occupations include government officials, cadres, and civil servants.

[b]Professional occupations include doctors, lawyers, journalists, teachers, etc.

[c]No profession includes temporary occupation or unemployed.

Out of the 501 valid responses, 216 (43.1%) were from male participants and the other 285 (56.9%) were from female participants. Further, the majority of respondents were between the ages of 20 and 39 years (276/501, 55.1%). Most of the respondents had earned a bachelor's degree (256/501, 51.1%), which indicated a high level of education among respondents. In terms of occupation, the largest group was students (104/501, 20.8%), followed by professionals (75/501, 15.0%) and enterprise managers (67/501, 13.4%).

Quality Control

We conducted quality control through the survey platform and via the investigators. Once on the platform, respondents were invited to fill in the questionnaire voluntarily. Each Internet Protocol (IP) address, computer, or username could only be used once. Also, there were various filtering rules for invalid answers, such as spending too little time on a questionnaire and trap rules to filter out random answers.

After the surveys were submitted, the respondents were screened by the investigators to retain the valid questionnaires. We excluded the following respondents: those who failed the attention check, where the answers to all the questions were the same or cyclical; those with response times of less than 120 seconds; those under 18 years old; and non-Chinese residents. Also, we checked the consistency between the IP address and the selected region, and questionnaires were eliminated if the IP addresses were not consistent.

Measures

The study instrument was modified from those in the relevant existing literature. Measurements and scales were translated into the appropriate Chinese versions to ensure the completeness and accuracy of instruments. After the repeated pretest, the final questionnaire was translated back into English, and the main semantics were not changed, indicating a strong correlation with the original English questionnaire (see Table 2 [21,29,34,44-52]). The instruments were measured using a 5-point Likert scale, ranging from 1 (highly disagree) to 5 (highly agree). All dimensions included three items, except the information characteristics construct, which included five items.

Table 2. Measurement items of the constructs.

Construct and variables	Measurement item
Information characteristics (IC) [21,45,46]	
IC1	COVID-19 prevention and treatment information on the internet is appropriate for my health demands.
IC2	COVID-19 prevention and treatment information on the internet is understandable.
IC3	COVID-19 prevention and treatment information on the internet is shared by most people (eg, by thumb-up or retweet).
IC4	The argument for COVID-19 prevention and treatment information on the internet is compelling.
IC5	The publisher of COVID-19 prevention and treatment information on the internet is experienced in the health field.
Social influence (SI) [44,47]	
SI1	People who are important to me think I should get COVID-19 prevention and treatment information from the internet.
SI2	My family and friends have obtained COVID-19 prevention and treatment information from the internet.
SI3	It is prevalent to get COVID-19 prevention and treatment information from the internet.
Perceived drawbacks (PD) [48,49]	
PD1	It may take me too much time or expense to adopt COVID-19 prevention and treatment information from the internet.
PD2	Adopting COVID-19 prevention and treatment information from the internet may cause psychological stress.
PD3	The health risks associated with the adoption of COVID-19 prevention and treatment information from the internet may outweigh the positive health outcomes.
Perceived benefits (PB) [29]	
PB1	It is important for me to adopt COVID-19 prevention and treatment information to reduce my risk of COVID-19 infection.
PB2	Adopting COVID-19 prevention and treatment information can help me stay healthy, which is very important to me.
PB3	The adoption of COVID-19 prevention and treatment information is valuable for me in order to adopt COVID-19 prevention behaviors.
Self-efficacy (SE) [50]	
SE1	I am confident that I can avoid COVID-19.
SE2	I can figure out how to avoid COVID-19 infection.
SE3	Even if I contract COVID-19, I can recover soon.
Perceived usefulness (PU) [46,51]	
PU1	The COVID-19 prevention and treatment information on the internet is valuable to me in preventing COVID-19.
PU2	I can make good use of COVID-19 prevention and treatment information on the internet in my life.
PU3	COVID-19 prevention and treatment information on the internet can improve the health of my family, friends, and myself.
Adoption intention (AI) [34,52]	
AI1	I will recommend this COVID-19 prevention and treatment information to my family and friends.
AI2	I will use this COVID-19 prevention and treatment information obtained from the internet in my daily life.
AI3	I would like to adopt COVID-19 prevention and treatment information, even if it takes my time or money (ie, to buy drugs, protective equipment, etc) to do so.

Ethics Approval and Consent to Participate

This study was approved in writing by the Medical Ethics Committee of Capital Medical University (No. Z2019SY014) and all participants gave informed consent.

Results

Overview

After data collection, the two-stage procedure of structural equation modeling (SEM) was applied to conduct data analysis [41]. The first procedure examined scale validity from the measurement model by confirmatory factor analysis (CFA), while the second procedure interpreted hypotheses testing by the structural model. Both SPSS Statistics for Windows, version 19.0 (IBM Corp), and SPSS Amos, version 24.0 (IBM Corp), were adopted as the tools for analyzing the data.

Measurement Model

Reliability

In this study, questionnaire items had a factor loading of 0.592 and above (see Table 3), which met the evaluation standard that the factor loading for construct measures must exceed 0.5 to be retained [53]. Cronbach α should be at least .70, and high reliability is assumed if it is greater than .80 [54]. The composite reliability (CR) value of greater than 0.70 represented high reliability [53]. All the constructs had both high Cronbach α and CR values, indicating high reliability (see Table 3).

Convergent Validity

Convergent validity measures the correlation of a dimension's multiple indicators. This study used the average variance extracted (AVE) to estimate the convergent validity [53]. A dimension with an AVE value over 0.50 would be considered as having high convergent validity [55]. As shown in Table 3, all dimensions had AVE values that were higher than the aforementioned cutoff values, which suggest good convergent validity.

In addition, all factor loadings for indicators measuring the same construct were statistically significant (see Table 3), suggesting that all indicators effectively measured their corresponding construct [56] and supported convergent validity.

Table 3. Reliability and convergent validity.

Construct and scale items	Factor loading[a]	Cronbach α[b]	Composite reliability coefficient[b]	Average variance extracted[b]
Information characteristics (IC)		.903	0.904	0.654
IC1	0.822	—	—	—
IC2	0.765	—	—	—
IC3	0.858	—	—	—
IC4	0.840	—	—	—
IC5	0.754	—	—	—
Social influence (SI)		.834	0.841	0.642
SI1	0.702	—	—	—
SI2	0.926	—	—	—
SI3	0.758	—	—	—
Perceived drawbacks (PD)		.753	0.756	0.510
PD1	0.633	—	—	—
PD2	0.789	—	—	—
PD3	0.712	—	—	—
Perceived benefits (PB)		.894	0.897	0.745
PB1	0.835	—	—	—
PB2	0.930	—	—	—
PB3	0.820	—	—	—
Self-efficacy (SE)		.773	0.788	0.558
SE1	0.837	—	—	—
SE2	0.790	—	—	—
SE3	0.592	—	—	—
Perceived usefulness (PU)		.884	0.885	0.719
PU1	0.826	—	—	—
PU2	0.869	—	—	—
PU3	0.849	—	—	—
Adoption intention (AI)		.868	0.875	0.702
AI1	0.831	—	—	—
AI2	0.944	—	—	—
AI3	0.723	—	—	—

[a]All factor loadings were significant at the $P<.001$ level.

[b]This value was calculated for each construct and not for individual items.

Discriminant Validity

Discriminant validity is achieved if the correlations between different constructs are relatively significant. The chi-square difference test can assess the discriminant validity of every two constructs by calculating the difference of the chi-square statistics for the constrained and unconstrained measurement models [57]. In this study, except for perceived drawbacks, the other six dimensions' chi-square difference tests were significant at the $P=.05$ level (see Table 4). Accordingly, the results demonstrated that discriminant validity was successfully achieved for the measurement model.

Table 4. Correlation analysis among constructs to determine discriminant validity.[a]

Construct	IC[b]	SI[c]	PD[d]	PB[e]	SE[f]	PU[g]	AI[h]
IC	0.809	—[i]	—	—	—	—	—
SI	0.729	0.801	—	—	—	—	—
PD	–0.044	–0.040	0.714	—	—	—	—
PB	0.741	0.799	–0.068	0.863	—	—	—
SE	0.437	0.421	0.023	0.512	0.747	—	—
PU	0.773	0.761	–0.152	0.857	0.555	0.848	—
AI	0.723	0.664	–0.038	0.741	0.479	0.792	0.838

[a]Diagonal elements are the square root of average variance extracted of the reflective scales. Off-diagonal elements are correlations between constructs.

[b]IC: information characteristics.

[c]SI: social influence.

[d]PD: perceived drawbacks.

[e]PB: perceived benefits.

[f]SE: self-efficacy.

[g]PU: perceived usefulness.

[h]AI: adoption intention.

[i]Repeated values were not included for easier comparison of table values.

Suppose the absolute value of the correlation coefficient is less than the square root of the AVE value. That would indicate that each construct has a certain correlation and a certain degree of differentiation between constructs, indicating that the scale data have an ideal discriminant validity. The value of the AVE square root of each construct was greater than the square of its correlation coefficient with the dimensions of all dimensions.

Structural Model Analysis

Based on the results of the CFA and modification index of indicator variables, six standard model fit criteria were used to assess the model's overall goodness of fit: ratio of the chi-square value to the degrees of freedom (χ^2/df), goodness-of-fit index (GFI), comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square residual, and root mean square error of approximation (RMSEA).

As shown in Table 5, comparison of all fit indices with their corresponding recommended values provided evidence of a good model fit: χ^2/df values were between 1.0 and 3.0; GFI, CFI, and TLI were all greater than 0.9; and RMSEA was smaller than 0.08. This demonstrated that the measurement model exhibited a tolerably good fit with the data collected [56].

Table 5. Goodness of fit of the measurement and structural models.

Statistical check	Goodness-of-fit criteria	Measurement model	Structural model	Result
χ^2/df	1.0-3.0	2.685	2.677	Good
Goodness-of-fit index	>0.9	0.908	0.908	Good
Comparative fit index	>0.9	0.953	0.953	Good
Tucker-Lewis index	>0.9	0.944	0.944	Good
Root mean square error of approximation	<0.08	0.058	0.058	Pass

Structure Model

Based on the results of SEM, the fit indices of the structural model are shown in Table 5. Under the same criteria, the structure model fits the observed data as well. Meanwhile, the estimated results of the structural model provided the path coefficients shown in Table 6. Among the eight hypotheses, six paths were supported based on the valid data, significant at the *P*=.01 level, while the remaining two paths were rejected according to SEM (ie, Hypothesis 2 and Hypothesis 5a).

Adoption intention was predicted by information characteristics (β=.266, *P*<.001) and perceived usefulness (β=.565, *P*<.001), which jointly explained 66.8% of the adoption intention variance. Information characteristics (β=.244, *P*<.001), perceived drawbacks (β=–.097, *P*=.002), perceived benefits (β=.512, *P*<.001), and self-efficacy (β=.141, *P*<.001) jointly determined perceived usefulness and explained 81.1% variance of perceived usefulness. In addition, perceived usefulness had the most significant influence on adoption intention. It revealed that perceived usefulness was an important indicator of adoption intention. Perceived benefits had the most significant direct influence on perceived usefulness, followed by information characteristics, while perceived drawbacks had a relatively low path coefficient, which indicated a negative effect at the same

time. Since the rejection of Hypothesis 5a, it means that perceived usefulness had a complete mediating effect on self-efficacy to adoption intention. This result confirmed perceived usefulness as an intermediary variable. Surprisingly, Hypothesis 2 was not supported in this study. The path coefficients supported six of all hypothesized relationships (see Table 6, Figure 2, and Multimedia Appendix 1).

Table 6. Testing results of the hypotheses.

Hypothesis	Path	Standardized path coefficient (β)	*P* value	Result
Hypothesis 1a	IC[a]→PU[b]	0.244	<.001	Supported
Hypothesis 1b	IC→AI[c]	0.266	<.001	Supported
Hypothesis 2	SI[d]→PU	0.115	.07	Rejected
Hypothesis 3	PD[e]→PU	–0.097	.002	Supported
Hypothesis 4	PB[f]→PU	0.512	<.001	Supported
Hypothesis 5a	SE[g]→PU	0.141	<.001	Supported
Hypothesis 5b	SE→AI	0.050	.25	Rejected
Hypothesis 6	PU→AI	0.565	<.001	Supported

[a]IC: information characteristics.

[b]PU: perceived usefulness.

[c]AI: adoption intention.

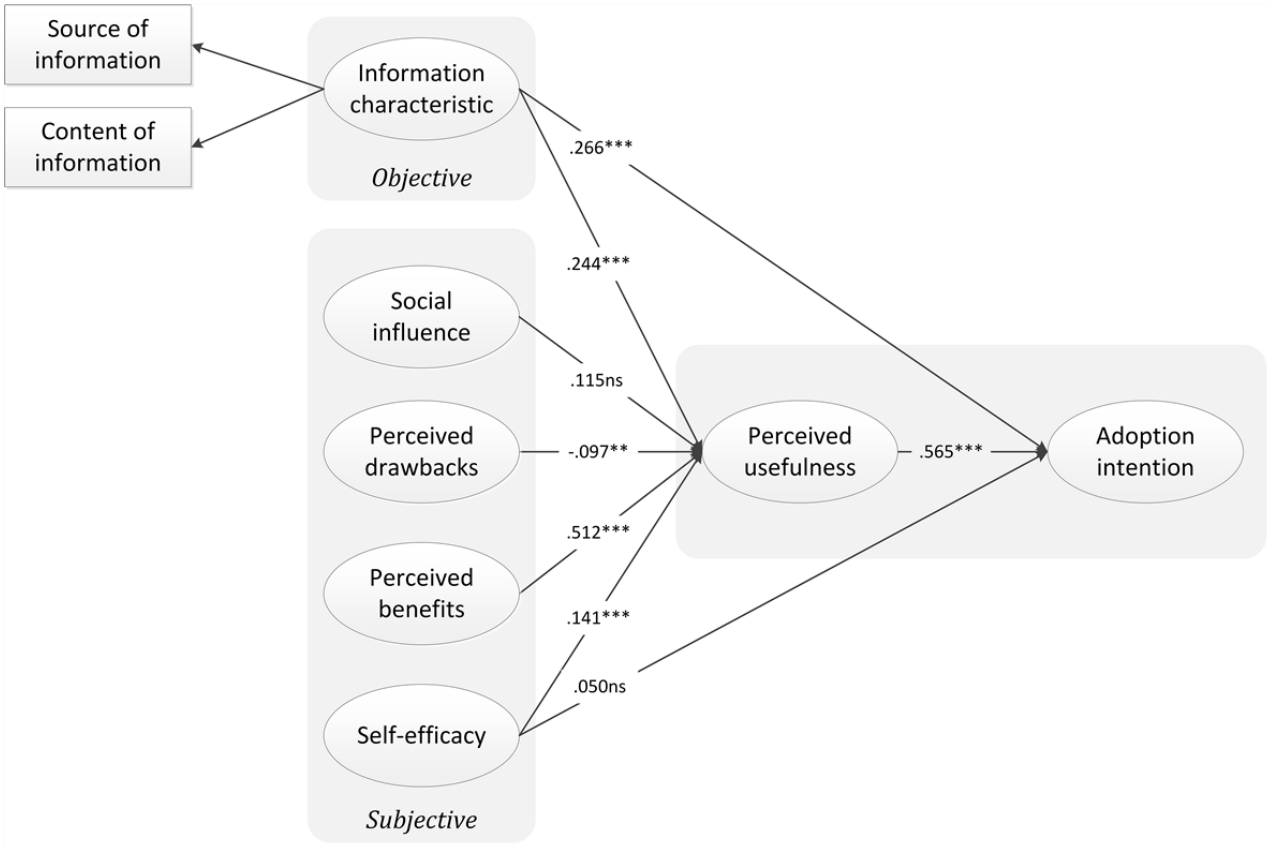
[d]SI: social influence.

[e]PD: perceived drawbacks.

[f]PB: perceived benefits.

[g]SE: self-efficacy.

Figure 2. Structural model results; standardized path coefficients (β) are shown. ***P<.001; **P<.01; ns: not significant (P>.05).



Analysis of Mediating Effects

To test the indirect effects, bias-corrected bootstrapping with 5000 iterations was implemented to obtain the structural model path significance levels for indirect effects [58]. Bootstrapping, a nonparametric approach, was superior to other approaches in testing mediation models because it does not assume multivariate normality [59,60]. Table 7 shows that perceived usefulness fully mediated the effect of perceived drawbacks, perceived benefits, and self-efficacy on adoption intention, whereas perceived usefulness partially mediated the effect of information characteristics on adoption intention.

Table 7. Mediation effect analysis.

Constructs of measurement	Standardized indirect effect	Bias-corrected values		Percentile		Mediating effect
		95% CI	*P* value	95% CI	*P* value	
IC[a]→PU[b]→AI[c]	0.138	0.069 to 0.218	<.001	0.071 to 0.220	<.001	Partial
SI[d]→PU→AI	0.065	–0.014 to 0.150	.10	–0.020 to 0.145	.13	No
PD[e]→PU→AI	–0.055	–0.094 to –0.023	.002	–0.091 to –0.020	.004	Fully
PB[f]→PU→AI	0.289	0.183 to 0.426	<.001	0.178 to 0.419	<.001	Fully
SE[g]→PU→AI	0.079	0.026 to 0.152	.002	0.026 to 0.152	.002	Fully

[a]IC: information characteristics.

[b]PU: perceived usefulness.

[c]AI: adoption intention.

[d]SI: social influence.

[e]PD: perceived drawbacks.

[f]PB: perceived benefits.

[g]SE: self-efficacy.

Discussion

Principal Findings

This study set out to determine the factors affecting people's intention of adopting health information on the internet, and the IAM-HBM model was applied. In general, this combined model provided an excellent fit to the data. Our research supports perceived drawbacks, perceived benefits, self-efficacy, perceived usefulness, and information characteristics as factors associated with the intention to adopt online prevention and treatment information to prevent an epidemic in the context of COVID-19.

Evidence-Based Information Plays an Important Role

In the study, information characteristics can strengthen perceived usefulness to adopt health information. In addition, the quality and the source of information influence the perceived usefulness and indirectly impact adoption intention. People are inclined to take in evidence-based information rather than misinformation on the internet [61,62]. Therefore, health communication should include more evidence-based information and should meet the public's health demand [7]. Swamped by information on the internet, if the information is expressed more reliably or if its publisher is more professional and authoritative, the public will have a more robust perception of the usefulness of the information, thus increasing individuals' willingness to adopt the information. At the same time, in the face of information overload, especially where some information constituted rumors, the public, who lack professional knowledge, will balance the outcomes of adoption information behavior. Many online repositories full of valuable content are underutilized, becoming "information junkyards" [63], and during the COVID-19 pandemic, an infodemic could be triggered by rumors. As information flow has improved, infodemic prevention and management using facts and evidence can mitigate the next infodemic [64,65].

Improving the Capacity to Obtain Public Information to Fight Mixed Messages

The public's perceived drawbacks, perceived benefits, and self-efficacy had significant influences on perceived usefulness. Members of the public conduct a cost-benefit analysis before adopting healthy behaviors, in which they weigh the effectiveness of the adoption against the possible cost and risk of time-consuming impediments. As a result, the more benefits and fewer drawbacks one perceives, the more that the health benefits of the adoption behavior outweigh the health risks, resulting in higher perceived usefulness by individuals in considering online health information to prevent COVID-19.

However, only by improving the public's media literacy and their ability to perceive information can they correctly recognize the obstacles and benefits. At the same time, greater health literacy can improve public health self-efficacy, resulting in increased confidence in adopting healthy behaviors or changing bad behaviors. Many people have limited health literacy [66]; health communication and education are the most cost-effective means to improve health literacy [67]. Therefore, in the release of COVID-19-related health information, attention should be paid to improving public information capacity. For example, officials could actively hold health lectures and disclose health information, and relevant experts could improve the public's capacity regarding obtaining information.

Simultaneously, with the continuous enrichment of social media, it is difficult for social communication based on social media to achieve full, comprehensive, and balanced transmission of information. We need to avoid falling into information cocoons and confirmation bias [68], and we need to measure the quality of information from an overall perspective.

Government Has Greater Influence Than Family and Friends

Surprisingly, our analysis did not support the hypothesis that social influence is positively associated with perceived usefulness of COVID-19 prevention and treatment information. This finding was counterintuitive, and previous research showed that social networks positively affect people by encouraging them to adopt different health behavior intentions [69].

We think that government press conferences and news-based public opinion during the COVID-19 pandemic in China have weakened social influence on people's perception and acceptance of health information for the following reasons. First, since the SARS epidemic in 2003, the Chinese government has reformed the news release concept and system. In response to the emergencies, the Chinese government issued a series of policy documents and established the State Council Information Office's three-level news spokesman system for all central ministries and provincial-level people's governments [70]. Second, governments at all levels use various channels to publicize, in a timely manner, the prevention and treatment information regarding COVID-19, the latest situation regarding the pandemic, and other public concerns, providing the public with a low threshold and low-cost direct information feedback channel [71]. The government also invited medical experts, such as Dr Zhong Nanshan, Head of the National Team for Control of Novel Coronavirus, to communicate with the public, and this strategy gained public trust [72]. Finally, China's political system practices high-quality and high-efficiency unified decision making, and they have strict controls over content such as social media [73]. For example, Facebook, Twitter, and YouTube are not allowed in mainland China [74], and information monitoring and timely rumor controls are also available within popular social media platforms, such as WeChat and Sina Weibo [75]. Therefore, the government exerted its full influence during the pandemic to position itself as the primary influence [76].

However, social media is both a source of the infodemic and a public health tool [77]. Therefore, it is necessary to include social media platforms in public information dissemination; rethinking the role of public communication will also be necessary to assume corresponding responsibility during the pandemic. The responsibility is not only to *delete* information but, as much as possible, to ensure the diversity of the information environment to a sufficient degree; this will enable high-quality public content and thereby increase public participation [64].

Conclusions

In a public health emergency, the online infodemic forces the public to negotiate with prevention and treatment information. By integrating IAM and HBM, this study provided the insight and understanding that perceived usefulness and adoption intention of online health information could be influenced by information characteristics, people's perceptions of drawbacks and benefits, and self-efficacy. Moreover, people also exhibit proactive behavior rather than reactive behavior. Thus, we should consider these factors to help the *informed public* obtain useful information via two approaches: one is to control the quality of information and the other is to improve the public's capacity to obtain information, in order to promote trusted information and to fight misinformation. This will, in turn, contribute to saving lives as the pandemic continues to unfold and run its course.

Limitations

We administered the questionnaire survey during the stage of the pandemic in mainland China when it was under control, which was when the outbreak in China had passed the initial panic stage. People at different stages of the pandemic may have been influenced differently by the influencing factors. Therefore, although we found that social influence had no significant effect on perceived usefulness of information, a more comprehensive future study is suggested to explore whether this is due to social context, stage of the pandemic, or other factors. Moreover, this cross-sectional study was conducted using the WJX web application, and sample populations had a certain amount of experience in filling out online questionnaires and internet use. Therefore, a more comprehensive future study is suggested to include offline and online participants to expand the framework's application scope.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (No. 71704118).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Path diagram of the information adoption model (IAM)–health belief model (HBM).
[PNG File , 444 KB - jmir_v23i3e23097_app1.png]

References

1. There are more than 2.36 million cases of COVID-19 globally. People's Daily. URL: https://news.sina.cn/gj/2020-04-20/detail-iirczymi7386584.d.html?vt=4&wm=3164_0008jishu [accessed 2020-04-20]
2. Munich Security Conference. World Health Organization. 2020 Feb 15. URL: https://www.who.int/director-general/speeches/detail/munich-security-conference [accessed 2020-02-15]
3. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for managing the COVID-19 infodemic: Methods and results of an online, crowdsourced WHO technical consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
4. Dumbrell D, Steele R. Putting the public into public health information dissemination: Social mediahealth-related web pages. In: Proceedings of the Seventeenth Australasian Document Computing Symposium. 2012 Dec Presented at: Seventeenth Australasian Document Computing Symposium; December 5-6, 2012; Dunedin New Zealand p. 135-138. [doi: 10.1145/2407085.2407104]
5. Risk A, Dzenowagis J. Review of internet health information quality initiatives. J Med Internet Res 2001;3(4):e28 [FREE Full text] [doi: 10.2196/jmir.3.4.e28] [Medline: 11772543]
6. Mheidly N, Fares J. Leveraging media and health communication strategies to overcome the COVID-19 infodemic. J Public Health Policy 2020 Dec;41(4):410-420 [FREE Full text] [doi: 10.1057/s41271-020-00247-w] [Medline: 32826935]
7. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb;395(10225):676. [doi: 10.1016/s0140-6736(20)30461-x]
8. Hou Z, Du F, Zhou X, Jiang H, Martin S, Larson H, et al. Cross-country comparison of public awareness, rumors, and behavioral responses to the COVID-19 epidemic: Infodemiology study. J Med Internet Res 2020 Aug 03;22(8):e21143 [FREE Full text] [doi: 10.2196/21143] [Medline: 32701460]
9. Larson HJ. A call to arms: Helping family, friends and communities navigate the COVID-19 infodemic. Nat Rev Immunol 2020 Aug;20(8):449-450 [FREE Full text] [doi: 10.1038/s41577-020-0380-8] [Medline: 32616908]
10. Ashrafi-Rizi H, Kazempour Z. Information diet in Covid-19 crisis: A commentary. Arch Acad Emerg Med 2020;8(1):e30 [FREE Full text] [Medline: 32232215]
11. Berg S. The well-informed patient: A new breed of health care consumer. Asthma Mag 2005 Jul;10(4):28-30. [doi: 10.1016/j.asthmamag.2005.06.005]
12. Yeh NC. Citizens' health information behaviors during SARS spread periods in Taiwan. J Libr Inf Stud 2003;1:95-110 [FREE Full text]
13. Teng Y, Bi D, Xie G, Jin Y, Huang Y, Lin B, et al. Dynamic forecasting of Zika epidemics using Google Trends. PLoS One 2017;12(1):e0165085 [FREE Full text] [doi: 10.1371/journal.pone.0165085] [Medline: 28060809]
14. Bora K, Das D, Barman B, Borah P. Are internet videos useful sources of information during global public health emergencies? A case study of YouTube videos during the 2015-16 Zika virus pandemic. Pathog Glob Health 2018 Sep;112(6):320-328 [FREE Full text] [doi: 10.1080/20477724.2018.1507784] [Medline: 30156974]
15. Hua J, Shaw R. Corona virus (COVID-19) "infodemic" and emerging issues through a data lens: The case of China. Int J Environ Res Public Health 2020 Mar 30;17(7):2309 [FREE Full text] [doi: 10.3390/ijerph17072309] [Medline: 32235433]
16. Li HO, Bailey A, Huynh D, Chan J. YouTube as a source of information on COVID-19: A pandemic of misinformation? BMJ Glob Health 2020 May;5(5):E002604 [FREE Full text] [doi: 10.1136/bmjgh-2020-002604] [Medline: 32409327]
17. Boyer C, Frossard C, Gaudinat A, Hanbury A, Falquetd G. How to sort trustworthy health online information? Improvements of the automated detection of HONcode criteria. Procedia Comput Sci 2017 Apr;121(4):940-949. [doi: 10.1016/j.procs.2017.11.122]
18. Houston TK, Sands DZ, Nash BR, Ford DE. Experiences of physicians who frequently use e-mail with patients. Health Commun 2003 Oct;15(4):515-525. [doi: 10.1207/s15327027hc1504_08]
19. Cailin O, James O. Why do we tend to believe rumors? Tencent. 2020 Feb 22. URL: https://new.qq.com/omn/20200222/20200222A0OKXT00.html [accessed 2020-02-23]
20. Rosenstock I, Strecher V, Becker M. The health belief model and HIV risk behavior change. In: DiClemente RJ, Peterson JL, editors. Preventing AIDS: Theories and Methods of Behavioral Interventions. Boston, MA: Springer; 1994:5-24.
21. Sussman SW, Siegal WS. Informational influence in organizations: An integrated approach to knowledge adoption. Inf Syst Res 2003 Mar;14(1):47-65. [doi: 10.1287/isre.14.1.47.14767]
22. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: A comparison of two theoretical models. Manage Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
23. Watts S, Zhang W. Capitalizing on content: Information adoption in two online communities. J Assoc Inf Syst 2008 Feb;9(2):73-94. [doi: 10.17705/1jais.00149]
24. Ha S, Ahn J. Why are you sharing others' tweets?: The impact of argument quality and source credibility on information sharing behavior. In: Proceedings of the 32nd International Conference on Information Systems. 2011 Presented at: 32nd International Conference on Information Systems; December 4-7, 2011; Shanghai, China p. 1195 URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1271&context=icis2011
25. Rosenstock IM. The health belief model and preventive health behavior. Health Educ Monogr 1974 Dec 01;2(4):354-386. [doi: 10.1177/109019817400200405]
26. Champion V, Skinner C. The health belief model. In: Glanz K, Rimer BK, Viswanath KV, editors. Health Behavior and Health Education: Theory, Research and Practice. 4th edition. San Francisco, CA: Jossey-Bass; 2008:46-65.

27. Harrison J, Mullen P, Green L. A meta-analysis of studies of the health belief model with adults. Health Educ Res 1992 Mar;7(1):107-116. [doi: 10.1093/her/7.1.107] [Medline: 10148735]
28. Dong Z. How to persuade adolescents to use nutrition labels: Effects of health consciousness, argument quality, and source credibility. Asian J Commun 2015 Feb 19;25(1):84-101. [doi: 10.1080/01292986.2014.989241]
29. Mou J, Shin D, Cohen J. Health beliefs and the valence framework in health information seeking behaviors. Inf Technol People 2016 Nov 07;29(4):876-900. [doi: 10.1108/itp-06-2015-0140]
30. Cheung CM, Lee MK, Rabjohn N. The impact of electronic word ‐ of ‐ mouth. Internet Res 2008 Jun 06;18(3):229-247. [doi: 10.1108/10662240810883290]
31. Erkan I, Evans C. The influence of eWOM in social media on consumers' purchase intentions: An extended approach to information adoption. Comput Human Behav 2016 Aug;61:47-55. [doi: 10.1016/j.chb.2016.03.003]
32. Ahadzadeh AS, Pahlevan Sharif S, Ong FS, Khong KW. Integrating health belief model and technology acceptance model: An investigation of health-related internet use. J Med Internet Res 2015 Feb 19;17(2):e45 [FREE Full text] [doi: 10.2196/jmir.3564] [Medline: 25700481]
33. Jeihooni AK, Rakhshani T. The effect of educational intervention based on health belief model and social support on promoting skin cancer preventive behaviors in a sample of Iranian farmers. J Cancer Educ 2019 Apr;34(2):392-401. [doi: 10.1007/s13187-017-1317-1] [Medline: 29313300]
34. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. MIS Q 2003 Jan 11;27(3):425-106. [doi: 10.2307/30036540]
35. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: Descriptive qualitative study. J Med Internet Res 2019 Apr 08;21(4):e11604 [FREE Full text] [doi: 10.2196/11604] [Medline: 30958272]
36. Redelmeier DA, Kraus NC. Patterns in patient access and utilization of online medical records: Analysis of MyChart. J Med Internet Res 2018 Feb 06;20(2):e43 [FREE Full text] [doi: 10.2196/jmir.8372] [Medline: 29410386]
37. Zeithaml VA. Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence. J Mark 2018 Nov 19;52(3):2-22. [doi: 10.1177/002224298805200302]
38. Kim H, Chan H, Gupta S. Value-based adoption of mobile internet: An empirical investigation. Decis Support Syst 2007 Feb;43(1):111-126. [doi: 10.1016/j.dss.2005.05.009]
39. Carpenter CJ. A meta-analysis of the effectiveness of health belief model variables in predicting behavior. Health Commun 2010 Dec;25(8):661-669. [doi: 10.1080/10410236.2010.521906] [Medline: 21153982]
40. Yun EK. Development and Testing of a Model of Consumers' Health Information Seeking Behavior on the Internet [doctoral thesis]. Seoul, South Korea: Seoul National University; 2008.
41. Qin Z, Huo C. Research on the continuous intention of health information seeking via social media: Bade on the view of revenue-expenditure. J Mod Inf 2020 May;40(05):66-77. [doi: 10.3969/j.issn.1008-0821.2020.05.009]
42. Rosenstock IM. Historical origins of the health belief model. Health Educ Monogr 1974 Dec 01;2(4):328-335. [doi: 10.1177/109019817400200403]
43. Aronson E, editor. The Social Animal. 8th edition. New York, NY: Worth Publishers; 1995.
44. Li C. Persuasive messages on information system acceptance: A theoretical extension of elaboration likelihood model and social influence theory. Comput Human Behav 2013 Jan;29(1):264-275. [doi: 10.1016/j.chb.2012.09.003]
45. Johnson J, Meischke H. A comprehensive model of cancer-related information seeking applied to magazines. Hum Commun Res 1993 Mar;19(3):343-367. [doi: 10.1111/j.1468-2958.1993.tb00305.x]
46. McKinney V, Yoon K, Zahedi F. The measurement of web-customer satisfaction: An expectation and disconfirmation approach. Inf Syst Res 2002 Sep;13(3):296-315. [doi: 10.1287/isre.13.3.296.76]
47. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]
48. Jacoby J, Kaplan L. The components of perceived risk: Advances in consumer research. In: Proceedings of the Third Annual Conference of the Association for Consumer Research. 1972 Presented at: Third Annual Conference of the Association for Consumer Research; November 3-5, 1972; Chicago, IL p. 382-393.
49. Peter JP, Tarpey Sr LX. A comparative analysis of three consumer decision strategies. J Consum Res 1975 Jun;2(1):29-37. [doi: 10.1086/208613]
50. Yoo W, Choi D. Predictors of expressing and receiving information on social networking sites during MERS-CoV outbreak in South Korea. J Risk Res 2019 Feb 11;23(7-8):912-927. [doi: 10.1080/13669877.2019.1569105]
51. Johnson JD, Donohue WA, Atkin CK, Johnson S. A comprehensive model of information seeking. Sci Commun 2016 Aug 18;16(3):274-303. [doi: 10.1177/1075547095016003003]
52. Tsai H, Bagozzi RP. Contribution behavior in virtual communities: Cogntive, emotional, and social influences. MIS Q 2014 Jan 1;38(1):143-163. [doi: 10.25300/misq/2014/38.1.07]
53. Hair J, Black W, Babin B, Anderson R. Multivariate Data Analysis: A Global Perspective. 7th edition. Upper Saddle River, NJ: Prentice Hall; 2010.
54. Nunnally JC, Bernstein IH. Psychometric Theory. 3rd edition. New York, NY: McGraw-Hill; 1994.

55. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
56. Anderson JC, Gerbing DW. Structural equation modeling in practice: A review and recommended two-step approach. Psychol Bull 1988 May;103(3):411-423. [doi: 10.1037/0033-2909.103.3.411]
57. Bagozzi RP, Phillips LW. Representing and testing organizational theories: A holistic construal. Adm Sci Q 1982 Sep;27(3):459-489. [doi: 10.2307/2392322]
58. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 2008 Aug;40(3):879-891. [doi: 10.3758/BRM.40.3.879]
59. Bollen KA, Stine R. Direct and indirect effects: Classical and bootstrap estimates of variability. Sociol Methodol 1990;20:115-140. [doi: 10.2307/271084]
60. Mackinnon DP, Lockwood CM, Williams J. Confidence limits for the indirect effect: Distribution of the product and resampling methods. Multivariate Behav Res 2004 Jan 01;39(1):99 [FREE Full text] [doi: 10.1207/s15327906mbr3901_4] [Medline: 20157642]
61. Pulido CM, Villarejo-Carballido B, Redondo-Sama G, Gómez A. COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information. Int Sociol 2020 Apr 15;35(4):377-392. [doi: 10.1177/0268580920914755]
62. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: A systematic review. JAMA 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
63. McDermott R. Why information technology inspired but cannot deliver knowledge management. Calif Manage Rev 1999 Jul;41(4):103-117. [doi: 10.2307/41166012]
64. Eysenbach G. How to fight an infodemic: The four pillars of infodemic management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
65. Guterres A. Tweet. Twitter. 2020 Apr 14. URL: https://twitter.com/antonioguterres/status/1250095790959267841?s=20 [accessed 2020-04-14]
66. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: An updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
67. Institute of Medicine. Health Literacy: A Prescription to End Confusion. Washington, DC: The National Academies Press; 2004.
68. Sunstein CR. Infotopia: How Many Minds Produce Knowledge. New York, NY: Oxford University Press; 2006.
69. Ryu I, Choi H. The effect of social influence on flow, perceived usefulness and intention to use in online community. J Inf Syst 2008 Jun 30;17(2):113-135. [doi: 10.5859/kais.2008.17.2.113]
70. Tian L, Yang Y. The quality assessment study of local government press conferences is based on case studies of COVID-19 press conferences in H province. Newsmaker 2020;5:8-13 [FREE Full text] [doi: 10.16017/j.cnki.xwahz.2020.05.003]
71. Zhong Nanshan's 12 latest judgments on the epidemic are very impressive. People's Daily. 2020 Apr 12. URL: https://baijiahao.baidu.com/s?id=1663732306125815219&wfr=spider&for=pc [accessed 2021-03-05]
72. Tencent News. Zhong Nanshan contributed to the coVID-19 outbreak?. URL: https://xw.qq.com/partner/standard/20201015A0DFVC/20201015A0DFVC00?ADTAG=standard&pgv_ref=standard [accessed 2020-10-17]
73. Fang N. The country's fight against the epidemic highlights the strength of China's political system. People's Daily. 2020 Jul 10. URL: http://theory.people.com.cn/n1/2020/0710/c40531-31778127.html [accessed 2020-07-10]
74. Sonmez F. How Chinese officials 'like' banned Facebook. Phys.org. 2014 May 18. URL: https://phys.org/news/2014-05-chinese-facebook.html [accessed 2014-05-18]
75. Here are the tools to check and verify rumors about the epidemic. Cyberspace Administration of China. 2020 Feb 18. URL: http://www.cac.gov.cn/2020-02/18/c_1583567100466237.htm [accessed 2020-02-18]
76. Public opinion leadership of new media from the perspective of COVID-19 epidemic prevention and control publicity. CN-Healthcare. 2020 Jul 28. URL: https://www.cn-healthcare.com/articlewm/20200727/content-1133260.html [accessed 2020-07-28]
77. Ahmed ST. Managing news overload (MNO): The COVID-19 infodemic. Information 2020 Jul 25;11(8):375. [doi: 10.3390/info11080375]

Abbreviations

AVE: average variance extracted
CFA: confirmatory factor analysis
CFI: comparative fit index
CR: composite reliability
ELM: elaboration likelihood model
GFI: goodness-of-fit index

HBM: health belief model
IACM: information acceptance model
IAM: information adoption model
IP: Internet Protocol
QR: Quick Response
RMSEA: root mean square error of approximation
SEM: structural equation modeling
TAM: technology acceptance model
TLI: Tucker-Lewis index
WHO: World Health Organization

Edited by C Basch; submitted 13.08.20; peer-reviewed by A Chang, J Manganello; comments to author 30.09.20; revised version received 20.10.20; accepted 17.02.21; published 11.03.21.

Please cite as:
Han Y, Jiang B, Guo R
Factors Affecting Public Adoption of COVID-19 Prevention and Treatment Information During an Infodemic: Cross-sectional Survey Study
J Med Internet Res 2021;23(3):e23097
URL: https://www.jmir.org/2021/3/e23097
doi:10.2196/23097
PMID:33600348

Original Paper

Antecedents of Individuals' Concerns Regarding Hospital Hygiene and Surgery Postponement During the COVID-19 Pandemic: Cross-sectional, Web-Based Survey Study

Thomas Ostermann[1], Prof Dr; Julia Gampe[1], BSc; Jan Philipp Röer[1], Prof Dr; Theda Radtke[2], Prof Dr

[1]Department of Psychology and Psychotherapy, Witten/Herdecke University, Witten, Germany

[2]Health Psychology and Applied Diagnostics, University of Wuppertal, Wuppertal, Germany

Corresponding Author:
Thomas Ostermann, Prof Dr
Department of Psychology and Psychotherapy
Witten/Herdecke University
Alfred-Herrhausen-Str 50
Witten, 58448
Germany
Phone: 49 2302926707
Email: thomas.ostermann@uni-wh.de

Abstract

Background: The COVID-19 pandemic poses a major challenge to people's everyday lives. In the context of hospitalization, the pandemic is expected to have a strong influence on affective reactions and preventive behaviors. Research is needed to develop evidence-driven strategies for coping with the challenges of the pandemic. Therefore, this survey study investigates the effects that personality traits, risk-taking behaviors, and anxiety have on medical service–related affective reactions and anticipated behaviors during the COVID-19 pandemic.

Objective: The aim of this study was to identify key factors that are associated with individuals' concerns about hygiene in hospitals and the postponement of surgeries.

Methods: We conducted a cross-sectional, web-based survey of 929 residents in Germany (women: 792/929, 85.3%; age: mean 35.2 years, SD 12.9 years). Hypotheses were tested by conducting a saturated path analysis.

Results: We found that anxiety had a direct effect on people's concerns about safety (β=−.12, 95% CI −.20 to −.05) and hygiene in hospitals (β=.16, 95% CI .08 to .23). Risk-taking behaviors and personality traits were not associated with concerns about safety and hygiene in hospitals or anticipated behaviors.

Conclusions: Our findings suggest that distinct interventions and information campaigns are not necessary for individuals with different personality traits or different levels of risk-taking behavior. However, we recommend that health care workers should carefully address anxiety when interacting with patients.

(J Med Internet Res 2021;23(3):e24804) doi:10.2196/24804

KEYWORDS

COVID-19; public health; medical investigations; surgery; hospitalization, medical practices

Introduction

In Germany, the first COVID-19 case was confirmed at the end of January 2020, and COVID-19 incidence rates rose in the following 3 months. In response, the Robert Koch Institute (ie, the German federal government agency and research institute responsible for disease control and prevention) and the Federal Centre for Health Education made the following recommendations to slow the interpersonal transmission of SARS-CoV-2: limit social contact, refrain from traveling unless absolutely necessary, work from home wherever possible, encourage the use of medical masks and gloves, and strengthen hand hygiene practices [1]. At the same time, the European Center for Disease Prevention and Control published a checklist to prepare hospitals for the reception and care of patients with COVID-19. This checklist included items that were related to hand hygiene, personal protective equipment, and the postponement of operations that were unrelated to COVID-19 [2]. However, the implementation of these regulations,

particularly those regarding the use of personal protective equipment during the initial weeks of the pandemic in Germany, was hindered by a lack of adequate medical masks and clothing [3]. Considering the fact that SARS-CoV-2 infection can result in severe illness and death, especially in people aged >65 years and those with defined risk factors (eg, high blood pressure, diabetes, chronic respiratory diseases, and cancer) [4], a lack of personal protective equipment in hospitals and inadequate medical practices can result in affective reactions (eg, worries and concerns) and anticipated behaviors (eg, the denial of important operations) among the general population [5,6].

An example of an affective reaction resulting from a concern about an impending or anticipated threat is worrying about the lack of personal protective equipment in hospitals. Various factors, such as sociodemographic characteristics and personal values, can be used to predict affective reactions [5,7]. Further, affective reactions like concern or worry positively relate to anxiety [8] and negatively relate to risk-taking behaviors [9]. In addition, personality traits (eg, neuroticism) are linked with affective reactions [10]. During the COVID-19 pandemic, it is necessary to investigate the possible antecedents of affective reactions that relate to hospital equipment and medical practices. Such information is necessary for training health care workers to develop psychological skills for helping patients who experience worry, anxiety, and other emotional problems. It is also necessary to investigate how the COVID-19 pandemic affects people's reactions when they or a person close to them needs to be hospitalized to undergo surgery for treating an illness [11,12]. Studies have shown that the COVID-19 pandemic poses a considerable challenge to routine medical services. For example, a study reported that patients prefer to postpone their operations until after the pandemic has completely passed due to the uncertain environment [12]. However, none of the studies that have been conducted during the pandemic have investigated psychological concepts that might influence individuals' concerns about hospital hygiene and the postponement of surgeries. Studies on treatment-related decisions have suggested that personality traits, risk-taking behaviors, and anxiety are important factors that affect people's decisions to avoid visiting a hospital or doctor [13,14].

Based on previous pandemics, it is known that segmenting the population into subgroups (ie, sociodemographic subgroups) is important for designing and delivering messages about health risks and health protection measures [15,16]. However, even though this might be a useful and effective method, these subgroups do not account for several important psychological factors, such as personality traits or anxiety. These factors might be crucial antecedents of affective reactions to public health messages. They might also influence people's health-related decisions. Specifically, these factors may directly affect anticipated behaviors that relate to people's decisions to postpone a nonurgent surgery [12,17]. Therefore, this survey study aims to identify the key factors that are associated with hospitalization-related and medical service–related affective reactions and anticipated behaviors during the COVID-19 pandemic.

We hypothesized that individuals with low levels of openness, high levels of conscientiousness, low levels of extraversion, low levels of agreeableness, high levels of neuroticism, low levels of risk-taking behavior, and high levels of anxiety would experience high levels of negative affective reactions and exhibit high levels of anticipated preventive behaviors in response to hospitalization and medical service provision.

Methods

Survey Summary

This cross-sectional, web-based survey study took place between March 19 and April 17, 2020. To ensure that our survey was highly visible to potential respondents, it was distributed via social media, email, direct communication methods, and advertisements in various digital communication channels. The recruitment of participants mainly took place at the Department of Psychology of Witten/Herdecke University. All participants were residents of Germany who were aged ≥16 years. All procedures in this study were performed in accordance with the ethical standards of the institutional review board of the Department of Psychology and Psychotherapy of Witten/Herdecke University and those of the American Psychology Association [18,19]. A letter of approval can be obtained from the first author.

Measures

Summary of Survey Instruments

Prior to the survey, we screened potentially eligible test instruments and scales to assess their suitability for answering the hypotheses. We selected validated scales (ie, whenever possible) for measuring the different survey constructs. We also developed new scales to measure the COVID-19–specific aspects of the survey, as no validated instruments were available at the time of the survey. The development of survey items was based on existing scales from other behavioral domains.

The following survey items, which were answered by using a visual analog scale that ranged from 0 (ie, not at all) to 100 (ie, absolutely), served as dependent variables: affective reactions and anticipated behaviors.

Affective Reactions

Affective reactions [20] were measured with two items for assessing concerns about hospital safety, hospital hygiene, and medical practices during the COVID-19 pandemic. After providing a short introduction to place the questions in the context of the COVID-19 pandemic, the following questions were asked: (1) "recently there have been supply bottlenecks of mouthguards, disinfectants or similar for hospitals and medical practices. Do you feel safe in places like this"; and (2) "how big is your concern that due to supply bottlenecks a proper hygiene cannot be ensured in hospitals or medical practices?"

Anticipated Behaviors

Anticipated behaviors were measured with two items for assessing people's decisions to postpone their own surgery or advise a person close to them against surgery during the pandemic. These items were in line with previous studies [11,12]. After providing a short introduction to place the questions in the context of the COVID-19 pandemic, the

following questions were asked: (1) "assuming you were about to have a non-urgent surgery - how likely would you be to postpone this surgery"; and (2) "suppose a person very close to you was about to have a non-urgent surgery, how likely is it that you would advise against having the surgery?"

The following survey items served as independent variables: personality, risk-taking behaviors, and anxiety.

Personality

People's personalities were measured with the Big Five Inventory (BFI)-10, which is the short version of the BFI-44 [21]. The BFI assesses the following five personality traits: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism (OCEAN). Openness to experience refers to whether people are inventive/curious or consistent/cautious. Conscientiousness refers to whether people are efficient/organized or easygoing/careless. Extraversion refers to whether people are outgoing/energetic or solitary/reserved. Agreeableness refers to whether people are friendly/compassionate or challenging/detached. Neuroticism refers to whether people are sensitive/nervous or secure/confident. Our psychometric analyses indicated that BFI-10 scores sufficiently correlated with BFI-44 scores. Based on the average correlation value (r=0.83), 70% of the variance in BFI-44 scores could be explained. After 6-8 weeks, the BFI-10 had an average retest reliability value of 0.75.

Risk-Taking Behaviors

Risk-taking behaviors were assessed with the readiness to take risk/search for competition scale of the Hamburger Personality Inventory (HPI), which includes 14 items that are evaluated with a 4-point Likert scale (eg, "Ultimately, I am also unstoppable by massive threats"). HPI item scores are added to calculate a risk-taking score [22]. With a Cronbach α value of .85, the HPI has high content and construct validity. The HPI has a test-retest reliability value of 0.86 after 18 months. Additionally, HPI scores positively correlate with autonomy orientations (r=0.48), revolutionary tendencies (r=0.53), conflict skills (r=0.53), and competitive attitudes (r=0.60). These scores also negatively correlate with harm avoidance tendencies (r=−0.78).

Anxiety

Anxiety was measured with the German version of the Spielberger State-Trait Anxiety Inventory (STAI), which is one of the most commonly used standard tools for measuring anxiety. In research, STAI scores also function as an indicator of distress. The state anxiety portion of the STAI consists of 20 items that are evaluated on a 4-point Likert scale (eg, "I feel worried"). All item scores are added to calculate a state anxiety score [23]. Higher STAI scores indicate greater anxiety/distress. The STAI has Cronbach α values that range between .90 and .94, which means that it has high content and construct validity. According to the original publication [23], the test-retest reliability coefficients of the STAI range between 0.65 to 0.75 (ie, within 2 months of completing the STAI). These coefficients remained stable in our psychometric analyses.

To assess whether people's risk of contracting COVID-19 and information-seeking behaviors (ie, those related to COVID-19) had an impact on their worries and anticipated behaviors, the following constructs were included in our analysis as covariates: risk profile and information-seeking behaviors.

Risk Profile

Risk profiles were adapted in accordance with previous studies [4,24]. Our survey included seven dichotomous items (ie, yes=1; no=0) that asked about risk factors for contracting COVID-19 (ie, age of >60 years, chronic lung disease, autoimmune disease, diabetes, kidney or liver diseases, cancer, immune deficiency, and the intake of immunosuppressive remedies). The sum of the item scores was used as a risk profile.

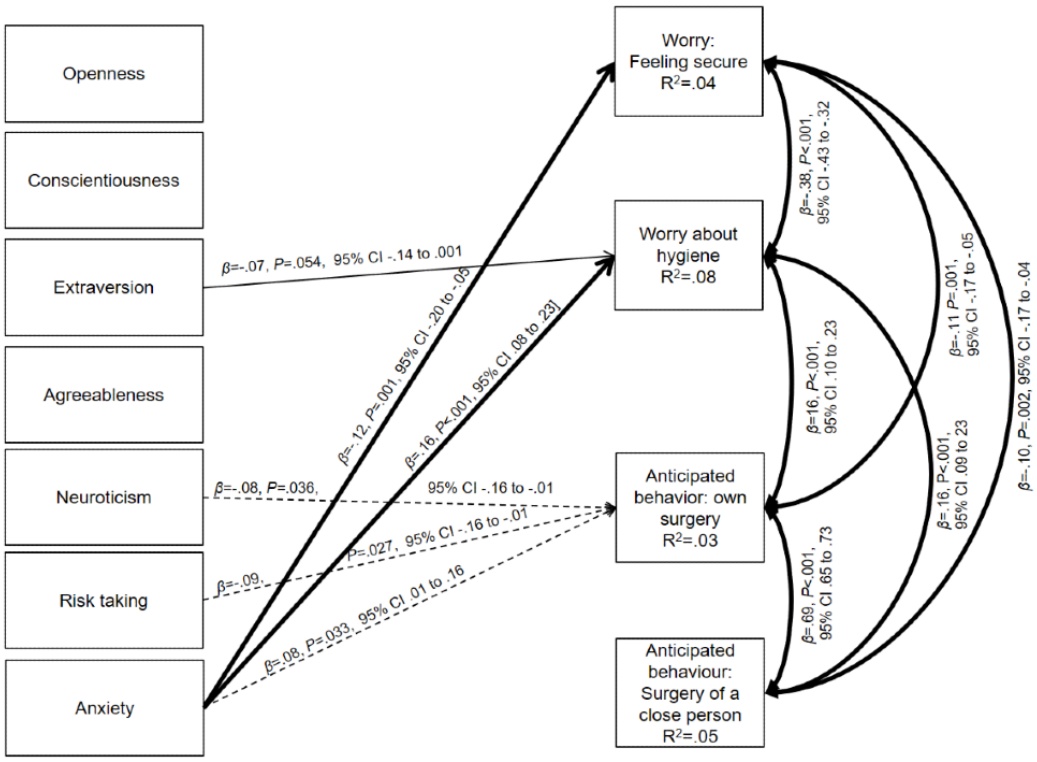
Information-Seeking Behaviors

Information-seeking behaviors were adapted in accordance with a previous study [25]. The behaviors we analyzed were in line with another study [26]. Our survey included six dichotomous items (ie, yes=1; no=0) that asked about the sources that people used to obtain information on COVID-19 (ie, television, internet blogs, social media, the website of the German federal government agency that is responsible for disease control and prevention, newspapers, and tabloid press articles). The sum of the item scores was used as an indicator of information-seeking behavior intensity. In addition, age, gender, and educational level (ie, a dichotomous variable that accounted for primary and secondary education) were introduced in the path model as covariates that needed to be controlled.

Statistical Analysis Strategy

Participants who fully completed the questionnaires were included in the statistical analysis. Descriptive statistical analyses were performed to describe the sample's characteristics in terms of the variables that were included in this study. In addition, bivariate correlation values were computed to examine associations among the variables. A saturated path model [27] with manifest variables was used to test whether OCEAN personality traits, anxiety, and risk-taking behaviors were related to worries about hospital safety and hygiene, worries about medical practices, and anticipated behaviors toward nonurgent surgeries (Figure 1). To assess whether people's risk of contracting COVID-19 and information-seeking behaviors (ie, those related to COVID-19) had an impact on their worries and anticipated behaviors, these variables were included in the analysis.

Figure 1. The hypothesized path model for identifying associations between independent variables (ie, personality traits, risk-taking behaviors, and anxiety) and dependent variables (ie, worries about safety, worries about hygiene, and anticipated behaviors). The model used data from 929 participants. We did not display the control variables (ie, risk profiles, information-seeking behaviors, age, gender, and education) to keep the model overview simple. Dotted lines refer to *P* values of ≥.01 and ≤.05. Bold lines refer to *P* values of <.001. Thin lines refer to *P* values of ≥.05. We did not display correlations between the control variables and outcomes to keep the model overview simple.



Age, gender, and educational level (ie, a dichotomous variable that accounted for primary and secondary education) were introduced in the model as covariates that needed to be controlled. All variables in the model were allowed to covary. Standardized regression coefficients (ie, β_i) for the path model (ie, the model for predicting affective reactions) and anticipated behaviors were calculated with the decomposition equation of correlations (ie, r_i), which is used to determine the direct and indirect effects that predictor variables (ie, X_i) have on dependent variables (ie, Y_j). In addition, the 95% CIs were calculated based on the 2.5 and 97.5 percentiles of the estimated SEs from bootstrapping [27]. As our study had four main dependent variables, Bonferroni correction was performed to lower the Cronbach α level for interpreting the results (ie, from .05 to .0125). To evaluate model fitness, the Chi-square test was used. According to Bollen and Long [28], Chi-square values should not be 2-5 times larger than the degrees of freedom. In addition, comparative fit index, Tucker-Lewis index, root mean square error of approximation, and standardized root mean square residual values were calculated as fit indices. Descriptive statistical analyses were conducted with SPSS, version 26 (IBM Corporation). The path analysis was conducted with Mplus, version 8.1 (Muthén & Muthén) [29].

Results

Descriptive Characteristics

Of the 1059 participants who took part in our survey, 929 (87.7%) had complete data sets. Thus, these 929 participants were included in the analyses. As indicated in Table 1, most of the participants (792/929, 85.3%) were female. The mean age of participants was 35.3 years (SD 12.9 years). Of the 929 participants, 890 (95.8%) stated that they were not infected with SARS-CoV-2, and only 7 (0.8%) stated that they were infected with SARS-CoV-2 (ie, at the time of the survey or before the survey). With respect to risk profiles, 683 (73.5%) participants reported that they did not exhibit any of the risk factors for contracting COVID-19, while 112 (12.1%) stated that they had a chronic lung disease. Almost all participants (884/929, 95.2%) subjectively felt restricted due to COVID-19–related regulations and measures. Details on participants' sociodemographic characteristics are provided in Table 1.

Descriptive statistics and correlations among the variables in the path model are reported in Tables 2 and 3. Worries about proper hospital hygiene and medical practices positively correlated with neuroticism (r=0.12) and anxiety (r=0.21). Further, all four dependent variables intercorrelated with each other. For example, worries about hygiene and worries about safety significantly correlated with each other (r=−0.40; P=.001).

Table 1. The sample's sociodemographic characteristics.

Sociodemographic variables	Value
Age (years), mean (SD)	35.3 (12.9)
Age (years), median (range)	32 (16-82)
Sex, n (%)	
Male	137 (14.7)
Female	792 (85.3)
Educational level, n (%)	
No school degree	3 (0.3)
Secondary school	6 (1.7)
Secondary modern education	108 (11.6)
Vocational baccalaureate	75 (8.1)
General baccalaureate	223 (24)
Applied science university diploma	116 (12.5)
Bachelor's degree	181 (19.5)
Master's degree	172 (18.5)
Doctorate degree or higher	34 (3.7)
COVID-19 status, n (%)	
Not infected	890 (95.8)
I was under suspicion	17 (1.8)
I am under suspicion	15 (1.6)
I was infected	5 (0.5)
I am infected	2 (0.2)
Risk profile, n (%)	
No risk factors	683 (73.5)
Aged >60 years	50 (5.4)
Chronic lung disease	112 (12.1)
Autoimmune disease	66 (7.1)
Diabetes	31 (3.3)
Cancer	15 (1.6)
Immunodeficiency	56 (6)
Intake of immunosuppressants	43 (4.6)
Information source, n (%)	
Friends and family	369 (39.7)
Television	553 (59.5)
Internet in general	401 (43.2)
Social media	402 (43.3)
Dedicated websites	758 (81.6)
Newspapers	495 (53.3)
Tabloid press articles	29 (3.1)
Feeling restricted due to COVID-19–related regulations and measures, n (%)	
Yes	884 (95.2)
No	45 (4.8)

Table 2. Bivariate correlations (ie, r values) among variables.

Variable	Conscientiousness	Extraversion	Agreeableness	Neuroticism	RTB[a]	Anxiety	Risk profile	Information profile	Feeling secure[b]	Hygiene[b]	Own surgery[c]	Surgery of a close person[c]
Openness												
r	0.03	0.03	0.03	−0.01	0.12[e]	0.01	0.02	−0.07[e]	0.01	0.03	0.01	<−0.01
P value	.39	.10	.35	.77	<.001	0.7	.57	0.05	.78	.38	.73	.91
Conscientiousness												
r	—[d]	0.12[e]	0.08[e]	−0.15[e]	0.12[e]	−0.12[e]	0.03	−0.08[e]	0.02	−0.04	0.02	0.06
P value	—	<.001	.02	<.001	<.001	<.001	.38	.02	.59	.25	.51	.09
Extraversion												
r	—	—	0.13[e]	−0.30[e]	0.22[e]	−0.22[e]	−0.01	<0.01	0.05	−0.11[e]	−0.01	−0.02
P value	—	—	<.001	<.001	<.001	<.001	.75	.90	.13	<.001	.88	.57
Agreeableness												
r	—	—	—	−0.12[e]	−0.04	−0.19[e]	−0.03	0.01	0.08[e]	-0.09[e]	−0.02	<−0.01
P value	—	—	—	<.001	.23	<.001	.41	.84	.02	.007	.47	.95
Neuroticism												
r	—	—	—	—	−0.32[e]	0.48[e]	0.03	0.03	−0.05	0.13[e]	−0.01	0.01
P value	—	—	—	—	<.001	<.001	.36	.43	.11	<.001	.69	.79
RTB												
r	—	—	—	—	—	−0.20[e]	0.04	−0.04	0.04	−0.05	−0.09[e]	−0.06
P value	—	—	—	—	—	<.001	.24	.23	.19	.14	.007	.07
Anxiety												
r	—	—	—	—	—	—	0.09[e]	0.07[e]	−0.14[e]	0.21[e]	0.08[e]	0.06
P value	—	—	—	—	—	—	.005	.04	<.001	<.001	.02	.05
Risk profile												
r	—	—	—	—	—	—	—	0.06	−0.07[e]	0.12[e]	0.02	0.07[e]
P value	—	—	—	—	—	—	—	.08	.03	<.001	.47	.04
Information profile												
r	—	—	—	—	—	—	—	—	<−0.01	0.04	−0.01	0.05
P value	—	—	—	—	—	—	—	—	.98	.25	.77	.16
Feeling secure[b]												
r	—	—	—	—	—	—	—	—	—	−0.40[e]	−0.13[e]	−0.13[e]
P value	—	—	—	—	—	—	—	—	—	<.001	<.001	<.001
Hygiene[b]												
r	—	—	—	—	—	—	—	—	—	—	0.18[e]	0.18[e]
P value	—	—	—	—	—	—	—	—	—	—	<.001	<.001
Own surgery[c]												
r	—	—	—	—	—	—	—	—	—	—	—	0.70[e]
P value	—	—	—	—	—	—	—	—	—	—	—	<.001

[a]RTB: risk-taking behavior.

[b]Refers to a worry.

[c]Refers to an anticipated behavior category.

[d]Not applicable.

[e]Significant at a level of $P<.05$.

Table 3. Mean and SD values of variables.

Variable	Value, mean (SD)
Openness	7.60 (2.02)
Conscientiousness	7.15 (1.65)
Extraversion	6.66 (1.97)
Agreeableness	6.20 (1.58)
Neuroticism	6.26 (2.04)
Risk-taking behavior	31.26 (7.13)
Anxiety	43.91 (12.23)
Risk profile	0.40 (0.80)
Information profile	3.24 (1.36)
Worries about feeling secure	48.32 (28.23)
Worries about hygiene	57.33 (30.35)
Anticipated behavior relating to own surgery	80.46 (28.45)
Anticipated behavior relating to the surgery of a close person	77.32 (28.93)

Results From the Path Model

The path model predicted the associations between independent variables (ie, personality, risk-taking behaviors, and anxiety) and dependent variables (ie, feelings about security, worries about hospital hygiene and medical practices, and anticipated behaviors that relate to people's decisions to postpone their own surgery or advise a person close to them against surgery). Figure 1 presents the parameter estimates of the model (ie, standardized solutions).

The following model-data fit indices were obtained: Chi-square value ($\chi^2_{54}=942.94$; N=929; $P<.001$), comparative fit index (1.00), Tucker-Lewis Index (1.00), root mean square error of approximation (<.01), and standardized root mean square residual (<.01). These values indicated a moderate model fitness. Table 4 provides the standardized regression coefficients of the path model, which was used to predict affective reactions and anticipated behaviors.

Table 4. Standardized regression coefficients of the path model, which was used to predict affective reactions and anticipated behaviors.

Path predictors	Affective reactions		Anticipated behaviors	
	Feeling secure, β (95% CI)	Concerns about hygiene, β (95% CI)	Own surgery, β (95% CI)	Surgery of a close person, β (95% CI)
Openness	<.01 (−.07 to .08)	.02 (−.05 to .09)	.02 (−.04 to .09)	−.02 (−.09 to .05)
Conscientiousness	.01 (−.05 to .08)	−.03[a] (−.09 to .04)	.01 (−.06 to .08)	.03 (−.03 to .10)
Extraversion	.03 (−.04 to .10)	−.07[a] (−.14 to <.01)	.01 (−.06 to .08)	−.02 (−.09 to .05)
Agreeableness	.05 (−.02 to .12)	−.05[a] (−.11 to .02)	−.03 (−.10 to .05)	<.01 (−.06 to .07)
Neuroticism	.05 (−.03 to .13)	.01 (−.07 to .09)	−.08[b] (−.16 to −.01)	−.02 (−.10 to .06)
Risk-taking behavior	.01 (−.07 to .08)	<.01 (−.07 to .08)	−.09[b] (−.16 to −.01)	−.04 (−.12 to .03)
Anxiety	−.12[c] (−.20 to −.05)	.16[d] (.08 to .23)	.08[b] (.01 to .16)	.05 (−.03 to .13)
Risk profile	−.06 (−.14 to .01)	.08[b] (.02 to .14)	<.01 (−.07 to .08)	.01 (−.05 to .08)
Information-seeking behavior	−.01 (−.09 to .07)	−.01 (−.07 to .05)	.01 (−.14 to .15)	.03 (−.05 to .10)
Gender[e]	.09[b] (.02 to .17)	−.06[a] (−.12 to .01)	−.08[a] (−.16 to .01)	−.08[b] (−.15 to −.01)
Age	−.05 (−.13 to .03)	.10[c] (.03 to .17)	.09[b] (.01 to .16)	.18[d] (.11 to .25)
Education[f]	.03 (−.05 to .11)	−.08[b] (−.16 to <−.01)	−.06 (−.16 to .04)	−.11[c] (−.19 to −.03)

[a]Significant at a level of $P<.10$.

[b]Significant at a level of $P<.05$.

[c]Significant at a level of $P<.001$.

[d]Significant at a level of $P<.001$.

[e]In the path model, women were given a value of 1 and men were given a value of 2.

[f]In the path model, secondary education was given a value of 1 and tertiary education was given a value of 2.

As outlined in Figure 1, the feeling of security with regard to hospitals and medical practices was significantly negatively related to anxiety (β=−.12; $P=.001$), which is in line with our hypothesis. Further, affective reactions to hospital hygiene and medical practices resulting from a bottleneck of appropriate personal protective equipment for health care workers were significantly positively related to anxiety (β=.16; $P<.001$) and nonsignificantly negatively related to extraversion (β=−.07; $P=.054$). Although anticipated behaviors that relate to advising a close person against surgery did not correlate with any of our hypothesized variables, anticipated behaviors that relate to one's own surgery were negatively associated with neuroticism (β=−.08; $P=.04$) and risk-taking behaviors (β=−.09; $P=.03$). Such anticipated behaviors were also positively associated with anxiety (β=.08; $P=.03$). All of these associations however were not statistically significant after the Bonferroni correction. No other associations between the independent and dependent variables were found. However, women and older participants reported that they experienced higher levels of negative affective reactions and anticipated behaviors compared to men and younger participants, respectively.

Discussion

Principal Findings

To the best of our knowledge, our study is the first to investigate predictors of affective reactions that relate to hospital safety, hospital hygiene, and medical practices during the COVID-19 pandemic. We are also the first to investigate anticipated behaviors that relate to people's decisions to postpone their surgery or advise a person close to them against surgery during the pandemic. Our findings are in line with those of a German-Austrian survey [30], which found that anxiety was positively related to security actions. Our results suggest that state anxiety is the most influential factor of anticipated health-related behaviors and concerns about safety or hygiene. Apart from state anxiety, none of the other hypothesized predictors (eg, risk-taking behaviors) or personality factors (eg, agreeableness or openness) had any significant association with affective reactions or anticipated behaviors. This is contradictory to the recent findings of Martin [31], who found that agreeableness was related to the perceived severity of the COVID-19 pandemic, and openness was related to low levels of anxiety with regard to contracting COVID-19. Although previous studies have suggested that individuals with high levels of neuroticism exhibit pronounced negative reactions to stressful events [32], our findings show that neuroticism was not associated with anticipated behaviors during the COVID-19 pandemic. In our study, we found that people with high levels of neuroticism were less likely to postpone their own surgery. This finding is comparable to that of an early US survey, which found that neuroticism was associated with high levels of concern [33]. The similarities in these results could be explained by the age of our participants. It is possible that our relatively

young participants were not able to accurately imagine a scenario in which they are hospitalized. This might have influenced participants' responses to our survey. With regard to the relationship between health policy formation and public responses [34], the most important finding of our study was that anxiety was related to both the affective reactions and anticipated behaviors of the participants. Allgleton and Kippax [35], who conducted an analysis on Australian HIV/AIDS policies, argued that suppressed anxiety can be used as a depressive position for eliciting a desired response in the general public [35]. Other studies [36-38] have also found that anxiety is an important predictive factor of taking preventive measures and exhibiting compliant behavioral responses during the 2009 influenza pandemic. Our findings support these empirical results. Health authorities should be aware that anxiety may not only affect individuals' behaviors but also the behaviors of organizations and systems (eg, splitting and blaming) [39]. In addition, anxiety resulting from the COVID-19 pandemic might also encourage individuals to consult a physician later than necessary (ie, to present their complaints). This has already occurred [40]. Such behavior may result in harms to health, the development of depression [41], or the chronification of disease.

Limitations and Implications for Future Research

Aside from the strengths of our study (eg, its large sample size), several limitations also need to be mentioned. First, due to the dynamic nature of the pandemic, we decided to use a random sample. However, due to our survey dissemination methods, our sample may not be representative of the German population. The generalizability of our results is open to empirical debate, as our sample mostly consisted of middle-aged and well-educated women. Research has shown that compared to men, women are more likely to actively seek health-related information and pay more attention to potential worldwide pandemics [42]. Second, our sample mainly consisted of middle-aged individuals. Therefore, it is reasonable to assume that older people are more likely to postpone surgeries and operations due to the COVID-19 pandemic, as they are more susceptible to the disease than middle-aged people. Older people also have stronger health care needs than middle-aged people [17]. Further research on the COVID-19–related concerns of older individuals is needed. Third, data collection took place during the beginning of the pandemic in Germany. Therefore, it remains unclear whether individuals would have the same affective reactions and anticipated behaviors later into the pandemic. Furthermore, the affective reactions and anticipated behaviors of people from urban areas should be distinguished from those of people from rural areas, as COVID-19 spreads at different rates in different geographical areas [43]. However, during the first phase of the pandemic in Germany, no considerable differences were found in infection and death rates [43]. This finding is also supported by the results of a recent survey study [44], wherein the authors did not find any substantial differences in behavioral intentions between participants from rural and urban regions in China. Fourth, our dependent variables were only measured with one item that used a visual analog scale. This was done to keep the survey concise and specific. Unfortunately, validated measures such as the COVID-19–Induced Anxiety Scale or the Protective Behaviors Towards COVID-19 Scale [45] were not available at the time of our survey. According to Heller et al [46] and Price et al [47], visual analog scales have sufficient psychometric measurement properties. Thus, they can be used when no validated instrument is available. Fifth, although the fitness of our path model was acceptable, it could have been better. However, it should be noted that as the sample size increases and the degrees of freedom remain constant, the Chi-square value increases. This leads to the problem of plausible models being rejected due to a significant Chi-square value. Therefore, too much emphasis should not be placed on the significance of the Chi-square statistic [48]. Furthermore, it should be noted that our data are cross-sectional in nature. As such, causal conclusions cannot be drawn from our data. Future studies should be longitudinal in nature.

Conclusions

Our results provide further insight into affective reactions and anticipated health-related behaviors during the COVID-19 pandemic. Our findings indicate that OCEAN personality traits are not associated with affective reactions and anticipated behaviors. Therefore, specific distinctions do not seem necessary when designing messages about health risks and health protection measures (ie, those related to hospital and medical practices during the COVID-19 pandemic). Even though future research is needed to confirm our results, health care workers should address the issues of patients with anxiety seriously and directly. Clear communication is necessary when providing information on the specific actions that hospitals and medical organizations perform to protect patients and health care workers. This could also help with preventing the cancellation of nonurgent surgeries in hospitals.

Authors' Contributions

TO, JR, and JG conceptualized this study. TO and TR designed the methodology of this study. JG, JR, and TO designed the survey. TO and TR performed the statistical analysis. TO, JG, and TR prepared the data. TO and TR wrote the initial manuscript draft. TO, JR, JG, and TR reviewed and edited the manuscript. TR created the figures and tables. TO supervised this study. All authors read and approved the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. RKI - Homepage. Robert Koch Institut. URL: https://www.rki.de/EN/Home/homepage_node.html [accessed 2021-02-23]
2. Checklist for hospitals preparing for the reception and care of coronavirus 2019 (COVID-19) patients. European Centre for Disease Prevention and Control. 2020 Feb. URL: https://www.ecdc.europa.eu/sites/default/files/documents/covid-19-checklist-hospitals-preparing-reception-care-coronavirus-patients.pdf [accessed 2021-02-23]
3. Boškoski I, Gallo C, Wallace MB, Costamagna G. COVID-19 pandemic and personal protective equipment shortage: protective efficacy comparing masks and scientific methods for respirator reuse. Gastrointest Endosc 2020 Sep;92(3):519-523 [FREE Full text] [doi: 10.1016/j.gie.2020.04.048] [Medline: 32353457]
4. Jordan RE, Adab P, Cheng KK. Covid-19: risk factors for severe disease and death. BMJ 2020 Mar 26;368:m1198. [doi: 10.1136/bmj.m1198] [Medline: 32217618]
5. Khosravi M. Perceived risk of COVID-19 pandemic: The role of public worry and trust. Electronic Journal of General Medicine 2020 Mar 27;17(4):em203 [FREE Full text] [doi: 10.29333/ejgm/7856]
6. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
7. Goodwin R, Gaines SO, Myers L, Neto F. Initial psychological responses to swine flu. Int J Behav Med 2011 Jun;18(2):88-92 [FREE Full text] [doi: 10.1007/s12529-010-9083-z] [Medline: 20195809]
8. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther 1990;28(6):487-495. [doi: 10.1016/0005-7967(90)90135-6] [Medline: 2076086]
9. Klein WMP, Zajac LE, Monin MM. Worry as a moderator of the association between risk perceptions and quitting intentions in young adult and adult smokers. Ann Behav Med 2009 Dec;38(3):256-261. [doi: 10.1007/s12160-009-9143-2] [Medline: 20049660]
10. Muris P, Roelofs J, Rassin E, Franken I, Mayer B. Mediating effects of rumination and worry on the links between neuroticism, anxiety and depression. Pers Individ Dif 2005 Oct;39(6):1105-1111. [doi: 10.1016/j.paid.2005.04.005]
11. Campi R, Tellini R, Grosso AA, Amparore D, Mari A, Viola L, et al. Deferring elective urologic surgery during the COVID-19 pandemic: The patients' perspective. Urology 2021 Jan;147:21-26 [FREE Full text] [doi: 10.1016/j.urology.2020.09.015] [Medline: 32979378]
12. Chang J, Wignadasan W, Kontoghiorghe C, Kayani B, Singh S, Plastow R, et al. Restarting elective orthopaedic services during the COVID-19 pandemic: Do patients want to have surgery? Bone Jt Open 2020 Nov 01;1(6):267-271 [FREE Full text] [doi: 10.1302/2046-3758.16.BJO-2020-0057] [Medline: 33215113]
13. Binder S, Nuscheler R. Risk-taking in vaccination, surgery, and gambling environments: Evidence from a framed laboratory experiment. Health Econ 2017 Dec;26 Suppl 3:76-96. [doi: 10.1002/hec.3620] [Medline: 29285871]
14. Orom H, Penner LA, West BT, Downs TM, Rayford W, Underwood W. Personality predicts prostate cancer treatment decision-making difficulty and satisfaction. Psychooncology 2009 Mar;18(3):290-299 [FREE Full text] [doi: 10.1002/pon.1385] [Medline: 18821530]
15. Holmes BJ. Communicating about emerging infectious disease: The importance of research. Health, Risk & Society 2008 Sep 05;10(4):349-360 [FREE Full text] [doi: 10.1080/13698570802166431]
16. Vaughan E, Tinker T. Effective health risk communication about pandemic influenza for vulnerable populations. Am J Public Health 2009 Oct;99 Suppl 2(Suppl 2):S324-S332. [doi: 10.2105/AJPH.2009.162537] [Medline: 19797744]
17. Clement ND, Oussedik S, Raza KI, Patton RFL, Smith K, Deehan DJ. The rate of patient deferral and barriers to going forward with elective orthopaedic surgery during the COVID-19 pandemic. Bone Jt Open 2020 Oct 21;1(10):663-668 [FREE Full text] [doi: 10.1302/2633-1462.110.BJO-2020-0135.R1] [Medline: 33215099]
18. American Psychological Association. Revision of Ethical Standard 3.04 of the "Ethical Principles of Psychologists and Code of Conduct" (2002, as amended 2010). Am Psychol 2016 Dec;71(9):900-900. [doi: 10.1037/amp0000102] [Medline: 28032786]
19. Chenneville T, Schwartz-Mette R. Ethical considerations for psychologists in the time of COVID-19. Am Psychol 2020;75(5):644-654. [doi: 10.1037/amp0000661] [Medline: 32437180]
20. Marinaci T, Carpinelli L, Venuleo C, Savarese G, Cavallo P. Emotional distress, psychosomatic symptoms and their relationship with institutional responses: A survey of Italian frontline medical staff during the Covid-19 pandemic. Heliyon 2020 Dec 16;6(12):e05766 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e05766] [Medline: 33364513]
21. Rammstedt B, John OP. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. J Res Pers 2007 Feb;41(1):203-212. [doi: 10.1016/j.jrp.2006.02.001]
22. Andresen B. HPI: Hamburger Persönlichkeitsinventar. Göttingen, Germany: Hogrefe; 2002.
23. Laux L, Glanzmann P, Schaffner P, Spielberger C. Das State-Trait-Angstinventar. Theoretische Grundlagen und Handanweisung. Weinheim: Beltz Test GmbH; 1981.
24. Li J, Huang DQ, Zou B, Yang H, Hui WZ, Rui F, et al. Epidemiology of COVID-19: A systematic review and meta-analysis of clinical characteristics, risk factors, and outcomes. J Med Virol 2021 Mar;93(3):1449-1458 [FREE Full text] [doi: 10.1002/jmv.26424] [Medline: 32790106]

25. Altizer KP, Grzywacz JG, Quandt SA, Bell R, Arcury TA. A qualitative analysis of how elders seek and disseminate health information. Gerontol Geriatr Educ 2014;35(4):337-353 [FREE Full text] [doi: 10.1080/02701960.2013.844693] [Medline: 24188253]
26. Moreno A, Fuentes-Lara C, Navarro C. Covid-19 communication management in Spain: Exploring the effect of information-seeking behavior and message reception in public's evaluation. El Profesional de la Información 2020 May 29;29(4):e290402 [FREE Full text] [doi: 10.3145/epi.2020.jul.02]
27. Wang J, Wang X. Structural Equation Modeling: Applications Using Mplus, Second Edition. Hoboken, New Jersey, United States: John Wiley & Sons Ltd; Oct 29, 2019.
28. Bollen KA, Long JS. Testing Structural Equation Models. Newbury Park: Sage Publications; Mar 1995.
29. Muthén BO, Muthén LK, Asparouhov T. Regression and Mediation Analysis Using Mplus. Los Angeles, CA: Muthén & Muthén; 2016.
30. Reiss S, Franchina V, Jutzi C, Willardt R, Jonas E. From anxiety to action-Experience of threat, emotional states, reactance, and action preferences in the early days of COVID-19 self-isolation in Germany and Austria. PLoS One 2020 Dec 8;15(12):e0243193. [doi: 10.1371/journal.pone.0243193] [Medline: 33290398]
31. Martin C. HEXACO traits, Big Five traits, and COVID-19. PsyArXiv. Preprint posted online on May 20, 2020 [FREE Full text] [doi: 10.31234/osf.io/c9gxe]
32. Mitchell LL, Zmora R, Finlay JM, Jutkowitz E, Gaugler JE. Do Big Five personality traits moderate the effects of stressful life events on health trajectories? Evidence from the health and retirement study. J Gerontol B Psychol Sci Soc Sci 2021 Jan 01;76(1):44-55. [doi: 10.1093/geronb/gbaa075] [Medline: 32478815]
33. Aschwanden D, Strickhouser JE, Sesker AA, Lee JH, Luchetti M, Stephan Y, et al. Psychological and behavioural responses to coronavirus disease 2019: The role of personality. Eur J Pers. Epub ahead of print 2020 Jul 08 [FREE Full text] [doi: 10.1002/per.2281] [Medline: 32836766]
34. Walsh K, Campbell S, Ashby M, Procter S. Public anxiety and health policy: A psychodynamic perspective. Soc Theory Health 2016 Oct 19;14:493-509 [FREE Full text] [doi: 10.1057/sth.2016.5]
35. Aggleton P, Kippax S. Australia's HIV-prevention response: Introduction to the special issue. AIDS Educ Prev 2014 Jun;26(3):187-190. [doi: 10.1521/aeap.2014.26.3.187] [Medline: 24846481]
36. Bults M, Beaujean DJ, de Zwart O, Kok G, van Empelen P, van Steenbergen JE, et al. Perceived risk, anxiety, and behavioural responses of the general public during the early phase of the Influenza A (H1N1) pandemic in the Netherlands: results of three consecutive online surveys. BMC Public Health 2011 Jan 03;11:2 [FREE Full text] [doi: 10.1186/1471-2458-11-2] [Medline: 21199571]
37. Jones JH, Salathé M. Early assessment of anxiety and behavioral response to novel swine-origin influenza A(H1N1). PLoS One 2009 Dec 03;4(12):e8032. [doi: 10.1371/journal.pone.0008032] [Medline: 19997505]
38. Rubin GJ, Amlôt R, Page L, Wessely S. Public perceptions, anxiety, and behaviour change in relation to the swine flu outbreak: cross sectional telephone survey. BMJ 2009 Jul 02;339:b2651 [FREE Full text] [doi: 10.1136/bmj.b2651] [Medline: 19574308]
39. Fotaki M, Hyde P. Organizational blind spots: Splitting, blame and idealization in the National Health Service. Hum Relat 2014 Jun 19;68(3):441-462. [doi: 10.1177/0018726714530012]
40. Bluro IM, Piñeiro DJ, Estada JLN. 'Cardiovascular Pandemic' in Argentina. Glob Heart 2020 Jul 30;15(1):51. [doi: 10.5334/gh.843] [Medline: 32923345]
41. Venkatesh A, Edirappuli S. Social distancing in covid-19: what are the mental health implications? BMJ 2020 Apr 06;369:m1379. [doi: 10.1136/bmj.m1379] [Medline: 32253182]
42. Ek S. Gender differences in health information behaviour: a Finnish population-based survey. Health Promot Int 2015 Sep;30(3):736-745. [doi: 10.1093/heapro/dat063] [Medline: 23985248]
43. Plümper T, Neumayer E. The pandemic predominantly hits poor neighbourhoods? SARS-CoV-2 infections and COVID-19 fatalities in German districts. Eur J Public Health 2020 Dec 11;30(6):1176-1180 [FREE Full text] [doi: 10.1093/eurpub/ckaa168] [Medline: 32818263]
44. Chen X, Chen H. Differences in preventive behaviors of COVID-19 between urban and rural residents: Lessons learned from a cross-sectional study in China. Int J Environ Res Public Health 2020 Jun 20;17(12):4437 [FREE Full text] [doi: 10.3390/ijerph17124437] [Medline: 32575700]
45. Riad A, Huang Y, Zheng L, Elavsky S. COVID-19 induced anxiety and protective behaviors during COVID-19 outbreak: Scale development and validation. medRxiv. Preprint posted online on May 9, 2020 [FREE Full text] [doi: 10.1101/2020.05.05.20050419]
46. Heller GX, Manuguerra M, Chow R. How to analyze the Visual Analogue Scale: Myths, truths and clinical relevance. Scand J Pain 2016 Oct;13:67-75. [doi: 10.1016/j.sjpain.2016.06.012] [Medline: 28850536]
47. Price DD, Staud R, Robinson ME. How should we use the visual analogue scale (VAS) in rehabilitation outcomes? II: Visual analogue scales as ratio scales: an alternative to the view of Kersten et al. J Rehabil Med 2012 Sep;44(9):800-801 [FREE Full text] [doi: 10.2340/16501977-1031] [Medline: 22915047]
48. Schermelleh-Engel K, Moosbruger H, Müller H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. Methods of Psychological Research 2003;8(2):23-74.

Abbreviations

BFI: Big Five Inventory
HPI: Hamburger Personality Inventory
OCEAN: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism
STAI: State-Trait Anxiety Inventory

Edited by G Eysenbach; submitted 06.10.20; peer-reviewed by Z Ren, P Nguyen, P Le-Mai; comments to author 22.10.20; revised version received 07.01.21; accepted 21.01.21; published 11.03.21.

Please cite as:
Ostermann T, Gampe J, Röer JP, Radtke T
Antecedents of Individuals' Concerns Regarding Hospital Hygiene and Surgery Postponement During the COVID-19 Pandemic: Cross-sectional, Web-Based Survey Study
J Med Internet Res 2021;23(3):e24804
URL: https://www.jmir.org/2021/3/e24804
doi:10.2196/24804
PMID:33617458

Original Paper

Examining Tweet Content and Engagement of Canadian Public Health Agencies and Decision Makers During COVID-19: Mixed Methods Analysis

Catherine E Slavik[1], MPH; Charlotte Buttle[1]; Shelby L Sturrock[2], MSc; J Connor Darlington[3], MSc; Niko Yiannakoulias[1], PhD

[1]School of Earth, Environment and Society, McMaster University, Hamilton, ON, Canada

[2]Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[3]School of Geography and Environmental Management, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Niko Yiannakoulias, PhD
School of Earth, Environment and Society
McMaster University
General Science Building
1280 Main Street West, Room 204
Hamilton, ON, L8S 4K1
Canada
Phone: 1 905 525 9140 ext 20118
Email: yiannan@mcmaster.ca

Abstract

Background: Effective communication during a health crisis can ease public concerns and promote the adoption of important risk-mitigating behaviors. Public health agencies and leaders have served as the primary communicators of information related to COVID-19, and a key part of their public outreach has taken place on social media platforms.

Objective: This study examined the content and engagement of COVID-19 tweets authored by Canadian public health agencies and decision makers. We propose ways for public health accounts to adjust their tweeting practices during public health crises to improve risk communication and maximize engagement.

Methods: We retrieved data from tweets by Canadian public health agencies and decision makers from January 1, 2020, to June 30, 2020. The Twitter accounts were categorized as belonging to either a public health agency, regional or local health department, provincial health authority, medical health officer, or minister of health. We analyzed trends in COVID-19 tweet engagement and conducted a content analysis on a stratified random sample of 485 tweets to examine the message functions and risk communication strategies used by each account type.

Results: We analyzed 32,737 tweets authored by 118 Canadian public health Twitter accounts, of which 6982 tweets were related to COVID-19. Medical health officers authored the largest percentage of COVID-19–related tweets (n=1337, 35%) relative to their total number of tweets and averaged the highest number of retweets per COVID-19 tweet (112 retweets per tweet). Public health agencies had the highest frequency of daily tweets about COVID-19 throughout the study period. Compared to tweets containing media and user mentions, hashtags and URLs were used in tweets more frequently by all account types, appearing in 69% (n=4798 tweets) and 68% (n=4781 tweets) of COVID-19–related tweets, respectively. Tweets containing hashtags also received the highest average retweets (47 retweets per tweet). Our content analysis revealed that of the three tweet message functions analyzed (information, action, community), tweets providing information were the most commonly used across most account types, constituting 39% (n=181) of all tweets; however, tweets promoting actions from users received higher than average retweets (55 retweets per tweet). When examining tweets that received one or more retweet (n=359), the difference between mean retweets across the message functions was statistically significant (P<.001). The risk communication strategies that we examined were not widely used by any account type, appearing in only 262 out of 485 tweets. However, when these strategies were used, these tweets received more retweets compared to tweets that did not use any risk communication strategies (P<.001) (61 retweets versus 13 retweets on average).

Conclusions: Public health agencies and decision makers should examine what messaging best meets the needs of their Twitter audiences to maximize sharing of their communications. Public health accounts that do not currently employ risk communication

strategies in their tweets may be missing an important opportunity to engage with users about the mitigation of health risks related to COVID-19.

(J Med Internet Res 2021;23(3):e24883) doi:10.2196/24883

KEYWORDS

COVID-19; coronavirus; pandemic; public health; Twitter; social media; engagement; risk communication; infodemiology; content analysis

Introduction

Background

On January 25, 2020, the first case of COVID-19 was reported in Canada in a man who had recently traveled to Wuhan, China, where the virus was first identified [1]. By mid-March, in the days after the World Health Organization (WHO) declared COVID-19 to be a pandemic, Canadian public health officials began to issue warnings against all nonessential travel, and soon local community transmission was confirmed to be the primary source of the transmission of cases in the country. By March 22, all Canadian provinces entered a state of emergency, ordered all nonessential businesses to close, and restricted public gatherings [2].

During this time, public health agencies and officials emerged as the de facto leaders and primary decision makers for setting evidence-based public health policies, practices, and norms. Daily updates from medical officers of health and other public health experts would set the course for how each jurisdiction would respond to COVID-19 and outline the public's role in "flattening the curve." Some early research suggests that Canadians listened to these messages and followed public health recommendations [3], and also stayed home, particularly in the early weeks of the pandemic, as demonstrated by decreases in the levels of people's movements tracked through Google's Community Mobility Reports [4].

Public Health Administration in Canada

The roles and responsibilities of Canadian public health institutions and individuals differ across the country and by levels of government. As a result, it is not always clear where the public should go to access and retrieve information during a health crisis. For example, the federal government's main role is to communicate national case numbers to all Canadians, coordinate control measures across provinces, and provide updates on national issues such as travel and the delivery of medical supplies. These activities are undertaken mostly by the Public Health Agency of Canada (PHAC), which was established as a separate agency of the federal department of health specifically to improve responses to infectious disease outbreaks after the severe acute respiratory syndrome (SARS) outbreak of 2003 [5].

Conversely, Canadian provinces are responsible for leading the emergency response, whereby ministries of health are tasked with communicating provincial updates on case counts, conducting surveillance and monitoring, providing guidance on infectious disease control measures and policies, and testing and screening practices [6]. Some provinces also operate regional and/or local-level public health units, which communicate information about local public health measures they have set and enforced based on provincial emergency orders. In addition to provincial health ministries, Ontario, British Columbia, and Quebec also have separate provincial-level public health agencies to provide scientific and technical advice on public health matters, conduct specialized data analytics, and provide updates on provincial testing capacity or other expert advice related to infectious disease control. Each of the levels of government described (federal, provincial, and regional/local where they exist) also appoints a medical officer of health to lead public health efforts in their respective jurisdiction, who often holds degrees and training in both medicine and public health.

The diversity in public health responses and responsibilities between institutions and individuals led virtually every Canadian province to take a different approach to crisis communications and information dissemination related to COVID-19. For example, provinces such as British Columbia have put their provincial medical officers of health in the spotlight, while others, including Ontario, have opted to have elected officials such as local premiers or health ministers lead some of the response. Occasionally this has resulted in contradictory messaging from multiple spokespeople, leaving the public confused and unsure whose guidance to follow [7]. This issue is particularly relevant on social media platforms like Twitter, where an abundance of information and misinformation has resulted in an infodemic [8], which can produce uncertainty and anxiety for individuals navigating an information overload [9]. Inconsistencies in health messaging can also erode public trust in the competence and credibility of public health agencies and decision makers [10].

Role of Public Health Agencies on Twitter

One of the ways in which the public has stayed informed on key information and updates on COVID-19 has been through the use of social media applications. Twitter in particular reported its biggest ever annual gain in daily users globally during the pandemic, which was up by 24% year over year during the first 3 months of 2020 [11]. On the one hand, an increase in Twitter users can lead to a more informed public as past research has suggested that a high proportion of users have identified Twitter as a major source of news for them [12]. On the other hand, an increase in Twitter users may also increase exposure to incorrect information or outright misinformation about COVID-19 [13]. Although the mass media have historically played a major role in the flow of information between public officials and the public during crises, the increased use of social media applications like Twitter has allowed members of the public to connect with governmental organizations and individuals more directly, largely

circumventing the need to follow other unofficial communicators [14].

During a health crisis such as a pandemic, the role of public health agencies and officials as communicators of timely and accurate information is especially crucial in helping the public form accurate perceptions of health risks and adapt their behaviors in ways that are necessary to mitigate risks [15]. In fact, social media platforms like Twitter have enabled users to seek and share information and news updates during past crises to help reduce feelings of uncertainty and cope with threats [16]. Some early studies of the COVID-19 pandemic have highlighted the need for public health officials to utilize more communication channels and exert their influence as risk communicators in a time when the global need for expert information and advice has peaked [17]. Information posted to social media especially at the early stages of any crisis or risk event tends to garner more traction online as users seek out updates. As such, it is critical for risk communicators to establish an early online presence and engage those users from the beginning [18].

One way to assess the online influence of Twitter accounts is to examine the engagement that their tweets receive. This can indicate how much an account's communications are being seen, studied, and shared. Tweet engagement can be indicated by various measures including the number of retweets (ie, shares of a tweet), likes (ie, number of times a user has seen and acknowledged or agreed with a tweet), or replies (ie, number of times a tweet has been commented on or responded to). Retweets in particular have been identified as an effective measure of engagement as they can indicate both the level of user agreement with a message and also the level of diffusion that message has undergone based on how many shares it has amassed from the original tweeter [19]. Beyond providing confirmation that some information or message has been disseminated to the public, quantifying tweet engagement based on retweets can provide a direct measure of the impact of that tweet on users. Some research has suggested that source credibility plays a role in garnering engagement; health agencies or individuals who appear to provide trustworthy information may be able to leverage their perceived legitimacy to gain more retweets and disseminate their information more broadly [20]. Researchers have also identified engagement strategies that can be used to increase user engagement to tweets. These strategies include the use of hashtags, URLs, user mentions (ie, direct mention of other Twitter user accounts), and media (eg, images or videos) [21].

Prior Work

Our study builds on past research that has examined the use of Twitter specifically by public health departments, agencies, and organizations. Most studies tend to focus on either examining the relationship between tweet features and levels of engagement and/or analyzing the content of the tweets to characterize the tweeting practices of particular accounts. For example, in a study of tweet engagement strategies used by 25 federal health agencies in the United States, it was found that hashtags, URLs, and user mentions were associated with an increased frequency of retweets [22]. In prior content analyses of tweets by state and/or local health departments in the United States, studies classifying the purpose of tweets (eg, whether tweets served to inform users or prompted them to perform some activity) have found that health departments mostly use Twitter to share health information [23-26]. However, other research has suggested that tweets whose function was to promote an action received more retweets than those with other functions [21]. Our work is also guided by research about prior pandemics such as tweeting trends during the H1N1 outbreak [27] and tweets covering Ebola health risks [28-30].

Previous work summarizing best practices in risk communication during broader risk events [31,32], as well as previous public health crises [33-35], underscore the importance of incorporating effective risk communication elements in messaging in order to reduce harm, clarify facts, and address public concerns. Beyond simply providing accurate descriptions of risks about the likelihood and consequences of harms, effective risk communication practices on Twitter may also include the use of messages promoting self-efficacy (ie, an individuals' beliefs that they have the ability to take action), providing reassurances, acknowledging concerns and uncertainties surrounding the situation, and indicating coordination of actions between experts. These strategies are viewed as important tools for organizations to augment their credibility and diffuse public fears [36]. Some literature has also noted the importance of applying strategic risk messaging across different outbreak phases by first focusing on information accuracy, then moving to reassurances to reduce uncertainty, and lastly, by emphasizing self-efficacy through individual actions and preventive measures [37].

Study Goal

The goal of our study was to characterize the content and level of engagement of COVID-19 tweets made by Canadian public health agencies and decision makers. Further, we propose recommendations for ways through which health agencies and decision makers could adjust their tweeting practices about COVID-19 and other future health crises to improve risk communication and maximize engagement. Our study seeks to answer four primary research questions (RQs):

- RQ1: which types of Canadian public health agencies and decision makers tweeted the most about COVID-19 and when?
- RQ2: how much engagement did tweets by Canadian public health agencies and decision makers receive during COVID-19? How did engagement change over time by account type?
- RQ3: did tweets containing Twitter engagement strategies receive more retweets than those that did not? How did the use of engagement strategies vary by account type?
- RQ4: did tweets from Canadian public health agencies and decision makers that employed a particular message function and risk communication strategy receive more retweets than others? How did the use of risk communication strategies in tweets from Canadian public health agencies and decision makers change over time?

Methods

Data Collection

A comprehensive list of Canadian public health institutions, agencies, and leaders was compiled after conducting a scoping review of provincial government websites. The resulting list of agencies and decision makers from this initial search was cross-referenced with the "Structural Profile of Public Health in Canada," a resource published by the National Collaborating Centre for Healthy Public Policy [38] that summarizes how public health is organized federally, provincially, and regionally across Canada. The names of each of these agencies and individuals were then manually searched using the Twitter interface to narrow the list to include only those that had a Twitter account (n=128). This list of agencies and decision makers was then used to pull tweets for the identified key players in Canadian public health communication on Twitter.

Twitter data were downloaded using the Twitter API accessed through R using the *rtweet* package (The R Foundation) [39]. An R script was created to go through the list of 128 identified Twitter accounts and download the most recent 3200 tweets of each account, which reflects the maximum number of tweets allowed for account-specific searches as imposed by the application programming interface for Twitter. Data were originally collected on May 23 but were recollected every day from June 1 to July 10 using an automated script that iteratively updated the number of interactions with past tweets (such as likes or retweets) and collected new tweets published during the month of June. The Twitter data set contained tweet-level data including the author's account name, Twitter handle, number of followers at the time of download; the date and time the tweet was published; whether the tweet was an original tweet or a retweet; the tweet's text, hashtags, user mentions, URLs, favorite and retweet count; and whether the tweet contained media (eg, image).

The Twitter data collected between May 23 and July 10 yielded 303,428 tweets from February 2010 to July 10, 2020. The data were then narrowed down to only include tweets authored between January 1, 2020, and June 30, 2020, resulting in 71,014 tweets. This time period was selected as China first reported the outbreak of the novel coronavirus to the WHO on January 1, 2020. Of the 128 Twitter accounts we sampled, 118 had authored tweets between this period. Finally, retweeted tweets were excluded, resulting in 45,310 tweets.

Most tweets in the sample were single standalone tweets within the tweet character limit, but some comprised longer tweet threads. In these cases, tweets were turned into threads using the following procedure. First, all tweets were sorted by publisher and date and time. Next, all tweets that were identified as self-replies (based on the "reply_to_screen_name" variable) were treated as part of a tweet thread. The start of a tweet thread was the tweet that immediately preceded the chain of tweets that were self-replies. This starting tweet, and all subsequent self-replies, were combined into a tweet thread. These tweets and tweet threads (n=32,737) will be referred to simply as tweets throughout the remainder of this paper. The tweets were then classified by whether they were about COVID-19 or not, based on whether they contained any of the following keywords: "covid*," "coronavirus," "ncov," "distanc*," "pand*," "tracing," "testandtrace," "curve," "stayhome," "handwashing," "mask," and "masque." These keywords were identified by scanning tweets within the sample and noting commonly used words in French- and English-language tweets describing COVID-19. During this scan, it was found that most French-language tweets about COVID-19 used #COVID19 to denote the tweet's topic relevant to COVID-19, which was a less common practice among English-language tweets. Hence, a larger number of English-language keywords were required to classify whether an English-language tweet was about COVID-19. This selection process resulted in 6982 tweets about COVID-19.

Additionally, we downloaded the publicly available COVID-19 data set from the Government of Canada [1] to plot active case counts over time alongside tweeting trends by public health accounts. The national COVID-19 data set aggregated case counts by the date that the case data were submitted to PHAC rather than the date that the cases were confirmed by the local health authority that collected the data.

Data Analysis

To classify the public health accounts by type, each account was categorized as belonging to either an agency or a decision maker. If the Twitter account belonged to an agency, it was classified as being one of 3 types: public health agency (ie, a federal or provincial public health agency, distinct from a health ministry due to its focus on public health), a regional or local health department (ie, a public health department that offers public health programs or services to communities at a scale smaller than the province), or a provincial health authority (ie, provincial health ministries or health authorities). If the account belonged to a decision maker, it was classified as being one of 2 types: medical officer of health (ie, the chief medical health officer of Canada and of each province, and regional or local medical health officers responsible for public health in smaller communities) or provincial minister of health (ie, elected government official who oversees health and public health agencies).

Tweet engagement was measured using retweet count. For tweet threads that contained multiple tweets, the maximum number of retweets obtained for any single tweet in the thread was used to measure engagement (rather than the average or total number of retweets for all tweets in the thread) to avoid double-counting or high-biased engagement.

Next, accounts were classified based on province, where applicable, or were otherwise identified as a "national" account (eg, PHAC and Canada's chief medical officer). This was done to select a stratified random sample of 501 tweets about COVID-19 for a qualitative content analysis (see Multimedia Appendix 1 for process flow chart). Since public health services and policies are primarily administered at the provincial level in Canada, we wanted our subset of tweets to capture enough geographic variation in the accounts across various provinces. We also wanted our subset of tweets to capture enough variation in tweeting by account type. Therefore, we used a stratified random sample with replacement using proportional weighting to randomly select tweets across various strata based on the

number of tweets each stratum contributed to the total sample. These comprised of 8 regional strata, which included British Columbia, Alberta, the Prairies (Saskatchewan and Manitoba), Ontario, Quebec, the Maritimes (Nova Scotia, New Brunswick, Prince Edward Island, and Newfoundland and Labrador), the Territories (Yukon and Northwest Territories), and Canada. No public health Twitter accounts from the Canadian territory of Nunavut were identified. Additionally, we randomly selected tweets across the two broad account types (ie, agencies and decision makers). The resulting stratified random sample of 501 tweets about COVID-19 contained 58 tweets from Canadian national-level accounts, 52 from Alberta, 66 from British Columbia, 50 from the Maritimes, 199 from Ontario, 47 from the Prairies, 17 from Quebec, and 12 from territorial accounts. The sample of 501 tweets about COVID-19 also contained 377 tweets from agencies and 124 tweets from decision makers.

Before beginning the content analysis, 3 researchers (CS, CB, SS) were trained on a set of 50 tweets randomly selected from the overall sample of 6982 COVID-19 tweets. This was done so that the researchers could familiarize themselves with the various variables for coding and troubleshoot any issues with the definitions of the coded variables. To distribute the 501 tweets for the content analysis equally among the three researchers, the French-language tweets (n=27) were first identified and allocated to the researcher with fluency in French. Among the remaining English-language tweets, 50 were randomly selected and allocated to each of the 3 researchers so that these overlapping tweets could be used to calculate the Krippendorff α value for interrater reliability. This resulted in one researcher coding 201 tweets (including the 27 French tweets), and the other two coding 200 tweets each. To integrate the 50 tweets that all 3 researchers had coded into the overall sample, one researcher's code was randomly selected from the 3 possible codes for each variable, such that the probability of selection was proportional to the frequency of that answer (eg, if two-thirds of coders agreed on a code, there was a 2 in 3 chance of that code being selected).

There were 10 coding variables in total. The first variable, media, captured the presence of media in the tweet and the type of media present (ie, image, video or document), if applicable. The next variable, message function, was coded using 3 mutually exclusive coding variables: information, action, or community. "Information" tweets included those whose main purpose was to inform, educate, or update the reader on case counts, disease transmission dynamics, policy changes, and COVID-19 symptoms. "Action" tweets included those whose main purpose was intended to prompt changes in the behaviors or actions of readers, such as encouraging social distancing, hygiene practices, or other harm-reducing behaviors. Finally, tweets were coded as "community" if their main purpose was community-building, identifying community supports and programs, or highlighting stories from or about the local community. Since threaded tweets could contain multiple message purposes, coding was based on the most prominent theme for the entire tweet thread. These variables for message function are consistent with those first proposed by Lovejoy and Saxton [40] for classifying the three main functions of organizations' Twitter use and have been used in similar research [21,23,25].

The final variable, use of a risk communication strategy, was coded using 6 nonmutually exclusive coding variables: corrective, risk, efficacy, concern, uncertainty, and experts. Tweets were classified as "corrective" if they corrected some incorrect information about COVID-19 or aimed to prevent the spread of misinformation. Tweets were classified as "risk" if they contained information that would help a reader make a judgment about the risk of contracting COVID-19 or experiencing health complications from COVID-19. This included tweets containing information regarding absolute risks, relative risks, as well as the identification of high-risk subpopulations. Tweets were classified as "efficacy" if they referenced an individual's or community's ability to execute an action or activity successfully resulting in some tangible benefit to health or a reduction of harm related to COVID-19. Tweets were classified as "concern" if they acknowledged the fears, concerns, worry, or anxiety associated with COVID-19. Tweets were classified as "uncertainty" if they acknowledged uncertainty, confusion, or a lack of available information about COVID-19. Finally, tweets were classified as "experts" if they implicitly or explicitly mentioned some agreement, coordination or collaboration between public health experts or other credible health organizations or individuals. The presence of any one of these 6 variables was used to indicate the use of any risk communication strategy in the tweet and were based in part on best practices in communication identified by Seeger [31] used to improve organizational and individual responses during crisis events (see Multimedia Appendix 2 for dataset of 501 manually coded Tweets).

Statistical Analysis

Each of the 3 coders worked independently through the same randomly selected 50 tweets, where each tweet had 10 variables to be codified. Krippendorff α [41] was used to measure the interrater agreement among coders, which was calculated using the R package *irr* [42]. Overall, interrater reliability was considered high (α=.829), with all 3 coders reporting total agreement on 453 out of the 500 answers (90.6%). Any codification of unstructured phenomena can have subjective biases, including when there is only one coder. However, the computed level of reliability suggests that there was largely internal agreement amongst the classification of variables within the sample such that results are less likely to be an artifact of internal disagreement or bias.

To assess whether differences in the mean number of retweets per tweet across each message function category were statistically significant, the nonparametric Kruskal-Wallis one-way ANOVA (analysis of variance) test was used since our data were not normally distributed.

We used the nonparametric Mann–Whitney–Wilcoxon Test to assess differences in the mean number of retweets per tweet between tweets containing at least one risk communication strategy and tweets containing no risk communication strategy.

Results

COVID-19 Tweets by Account Type

Our sample comprised 32,737 tweets, which included individual tweets and threads authored by 118 Canadian public health Twitter accounts (agencies and decision makers) between January 1, 2020, to June 30, 2020. Approximately 21% (n=6982) of all tweets contained content about COVID-19. Table 1 summarizes the characteristics of tweets in our sample by account type. Medical officers of health authored the largest percentage of tweets about COVID-19 relative to their total tweets (n=1337, 35%), representing the largest contribution of any account type. Conversely, accounts that belonged to provincial health ministers authored the smallest percentage of their tweets about COVID-19 (n=350, 18%). Accounts corresponding to Canadian medical officers of health also had the highest average number of retweets for COVID-19–related tweets, as well as the largest total follower count (summed across accounts in this category) at 416,611 total users (range 213–206,288 followers).

Table 1. Number of tweets and follower counts by account type, January 1, 2020, to June 30, 2020.

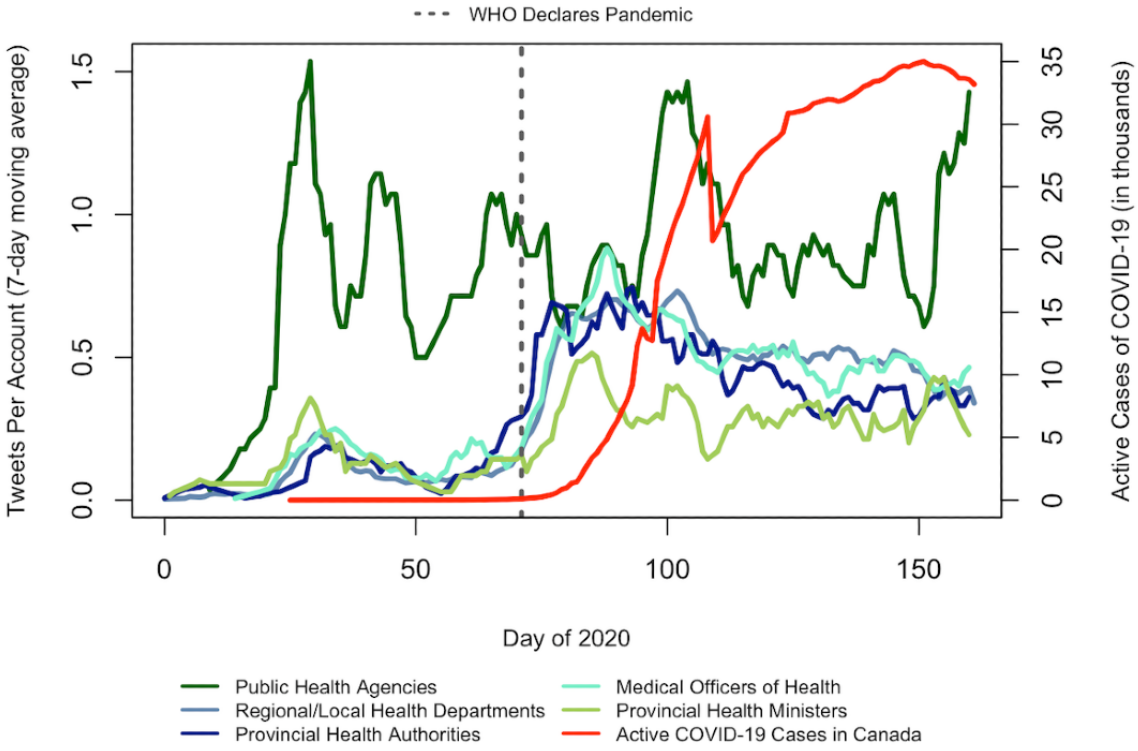
Account type	Twitter accounts, n	Total tweets[a], n	Mean tweets per account	Tweets about COVID-19, n (%)	Mean retweets per tweet about COVID-19	Total follower count[b], n (range)
Public health agencies	4	2272	568	524 (23)	60	407,546 (10,201-325,112)
Regional and local health departments	69	19,919	289	3832 (19)	10	406,108 (194-82,347)
Provincial health authorities	15	4778	319	939 (20)	13	170,387 (23-41,779)
Medical officers of health	22	3859	175	1337 (35)	112	416,611 (213-206,288)
Provincial health ministers	8	1909	239	350 (18)	52	134,019 (908-53,325)

[a]Equals the number of tweets and/or tweet threads authored by Twitter accounts per type between January 1 and June 30, 2020.

[b]Corresponds to the number of followers the accounts had at the end of the study period on June 30, 2020.

Figure 1 displays longitudinal trends in the daily rate of COVID-19 tweets stratified by account type using a 7-day moving average. All account types had an increase in their daily rate of COVID-19–related tweets after January 25 (day 25), when Canada's first case of COVID-19 was reported (Figure 1). Compared to the other account types, public health agencies authored the highest number of COVID-19–related tweets per account through most of the period studied. An exception to this trend was observed shortly after the WHO declared COVID-19 a pandemic, when the other account types increased their frequency of tweets about the disease. In the time between the WHO's declaration (day 70) and when Canada's COVID-19 cases first peaked (day 108), the daily number of COVID-19 tweets per account for regional and local health departments, provincial health authorities, and medical health officers appeared to converge (Figure 1).

Figure 1. Daily rate of COVID-19 tweets by account type, January 1, 2020, to June 30, 2020. WHO: World Health Organization.



Engagement

Figure 2 displays trends in the average number of retweets for COVID-19–related tweets over time by account type. Around day 28, 3 days after the first case of COVID-19 was confirmed in Canada, there was a large spike in the average number of retweets for public health agencies that lasted a few days before returning to baseline. For all account types, the next period of increase in retweets per tweet occurred around the time of the WHO's pandemic declaration (day 70). However, a few weeks after the pandemic was declared, retweets appeared to trend downward, even before COVID-19 cases peaked in Canada (Figure 2). The maximum daily average in retweets (381 retweets per tweet) was seen on day 80 among accounts belonging to medical officers of health; importantly, this was the day that Canada announced it would be closing its border to most noncitizens and nonpermanent residents. In contrast to other account types, medical officers of health maintained relatively high engagement (average of 50 or more retweets per COVID-19–related tweet) for a sustained period, beginning shortly before the WHO's pandemic declaration on day 70 and lasting until the second peak in COVID-19 cases in Canada on day 150. For accounts corresponding to provincial health ministers, daily retweets peaked on the same day as they did for medical health officers (day 80) but trended downwards shortly thereafter (Figure 2). Trends in retweets over time were similar for provincial health authorities and regional or local health departments (Figure 2).

Figure 2. Daily retweets per COVID-19 tweet by account type, January 1, 2020, to June 30, 2020. Avg: average; WHO: World Health Organization.

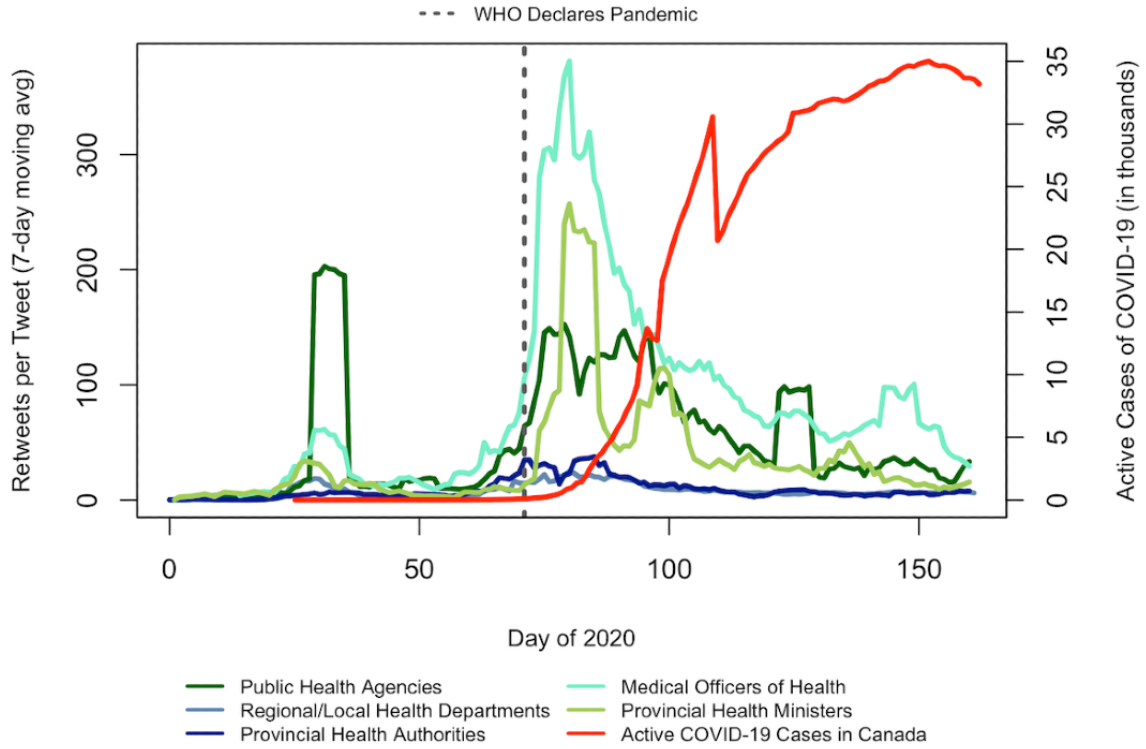


Table 2 summarizes the total number of tweets containing each engagement strategy, stratified by account type, as well as the sum and mean number of retweets by Twitter engagement strategy. User mentions were used less frequently than other engagement strategies and received the lowest mean retweets across all accounts. The most frequently used engagement strategies across all account types were hashtags (n=4798) and URLs (n=4781). These two engagement strategies appeared equally as frequently in tweets authored by public health agencies; however, the use of media (eg, images and videos) was associated with the highest average retweet count (67 retweets per tweet) for this account type. Similar findings were observed for provincial health authorities, which also used URLs most frequently (n=669) and received the highest average retweets on tweets containing media (18 retweets per tweet). Although accounts corresponding to regional and local health departments also incorporated URLs in more tweets than the other strategies (n=2766), they received the highest average retweets for tweets that contained hashtags (13 retweets per tweet). Tweets by regional and local health departments on average received fewer retweets per tweet compared to the other account types. Tweets by medical health officers that contained hashtags on average received the highest number of retweets per tweet compared to any engagement strategy and any account (134 retweets per tweet). Both medical officers of health and provincial health ministers used tweets containing media in fewer than half of their tweets, while the other three account types used media in more than half of their tweets.

Table 2. Summed tweet (n=6982) and retweet frequencies and percentages by engagement strategy and account type, January 1, 2020, to June 30, 2020.

Account type and engagement strategy	Number of tweets[a] containing each engagement strategy, n (%)	Summed retweets for all tweets containing each strategy, n	Mean retweets per tweet containing each strategy
Public health agencies			
Media	350 (67)	23,291	67
Hashtags	451 (86)	29,418	65
URLs	451 (86)	24,788	55
User mentions	139 (27)	2654	19
Regional and local health departments			
Media	2640 (69)	30,223	11
Hashtags	2462 (64)	30,824	13
URLs	2766 (72)	27,136	10
User mentions	753 (20)	4648	6
Provincial health authorities			
Media	477 (51)	8555	18
Hashtags	614 (65)	8333	14
URLs	669 (71)	6671	10
User mentions	326 (35)	1644	5
Medical officers of health			
Media	342 (26)	25,924	76
Hashtags	1058 (79)	141,841	134
URLs	681 (51)	61,697	91
User mentions	396 (30)	15,098	38
Provincial health ministers			
Media	157 (45)	5015	32
Hashtags	213 (61)	14,261	67
URLs	214 (61)	8134	38
User mentions	135 (39)	3756	28

[a]Equals the number of tweets and/or tweet threads authored by Twitter accounts per type between January 1 and June 30, 2020.

Content Analysis

During the content analysis of 501 tweets, 16 tweets were identified by the 3 researchers as not being directly related to content about COVID-19, resulting in a data set of 485 tweets about COVID-19. When coding the tweets for message function, 21 tweets were found to not have a classifiable purpose and were omitted from this part of the analysis. Table 3 summarizes the frequency and percentage of tweets identified as information, action, and community for each of the five account types in the sample. More than half of all coded tweets authored by public health agencies were classified as information tweets (n=17, 52%), which received the highest average number of retweets per tweet (56 retweets per tweet) compared to action (43 retweets per tweet) and community tweets (12 retweets per tweet) for these accounts. In our sample, tweets authored by regional and local health departments were most often classified as action tweets (n=101, 47%), and on average these received the highest number of retweets per tweet for this account type (10 retweets per tweet). Tweets corresponding to provincial health authorities, medical health officers, and provincial health ministers were mostly classified as information (n=47, 47%; n=56, 58%; and n=10, 53%, respectively); however, action tweets authored by each of these account types received more retweets per tweet on average compared to their information tweets (12, 259, and 44 retweets per tweet, respectively). The difference in mean retweets across the three message functions was not statistically significant (P=.18). However, when examining only those tweets that received one or more retweet (n=359), the difference between mean retweets across the three message functions was statistically significant (P<.001).

Table 3. Summed tweet (n=464) and retweet frequencies and percentages by message function and account type, January 1, 2020, to June 30, 2020.

Account type and message function	Number of tweets[a] with each message function, n (%)	Summed retweets for all tweets of each function, n	Mean retweets per tweet of each function
Public health agencies			
Information	17 (52)	957	56
Action	13 (39)	554	43
Community	3 (9)	36	12
Regional and local health departments			
Information	51 (24)	325	6
Action	101 (47)	1043	10
Community	64 (30)	223	3
Provincial health authorities			
Information	47 (47)	393	7
Action	33 (33)	388	12
Community	19 (19)	142	7
Medical officers of health			
Information	56 (58)	6172	103
Action	30 (31)	7765	259
Community	11 (11)	223	20
Provincial health ministers			
Information	10 (53)	293	27
Action	5 (26)	221	44
Community	4 (21)	90	23

[a]Equals the number of tweets and/or tweet threads authored by Twitter accounts per type between January 1 and June 30, 2020.

Table 4 summarizes the frequencies and percentages of risk communication strategies used by account type for the stratified random sample of COVID-19 tweets that were coded during the content analysis. Overall, the risk communication strategies that we examined were not very widely used and appeared only in 262 tweets out of our sample of 485 tweets. Since some tweets in our sample contained more than one type of risk communication strategy, as a result, there were 334 strategies used across the 262 tweets. Efficacy statements were the most commonly used strategy (efficacy accounted for more than one-third of the strategies used by each account type), and this strategy appeared in 163 of 262 tweets containing any risk communication strategy. For accounts corresponding to public health agencies, efficacy (n=12, 46%) and risk (n=8, 31%) statements were the most frequently used strategy; however, tweets containing these strategies were not the most retweeted. Instead, a single tweet thread containing corrective information that was authored by PHAC (which addressed misinformation on COVID-19) received the most retweets (134 retweets per tweet). Among regional and local health departments, the second most frequently used risk communication strategy after efficacy was addressing concern about COVID-19; however, their tweets containing corrective information received the most retweets on average (30 retweets per tweet). Provincial health authorities and medical health officers used risk statements at similar frequencies (n=12, 23% and n=24, 22%, respectively); however, tweets containing this strategy had fewer retweets on average among both account types when compared to tweets containing other risk communication strategies. Although medical health officers only authored 4 tweets containing statements that acknowledged the uncertainty around COVID-19, these tweets received the highest number of retweets per tweet compared to the other strategies used by that account type (358 retweets per tweet).

Table 4. Summed risk communication strategies (n=334) and percentages and retweet frequencies by strategy and account type, January 1, 2020, to June 30, 2020.

Account type and risk communication strategy[a]	Number of risk communication strategies[b] used by account type, n (%)	Summed retweets for all tweets containing each risk communication strategy, n	Mean retweets per tweet containing each strategy
Public health agencies			
Corrective	1 (4)	134	134
Risk	8 (31)	112	14
Efficacy	12 (46)	845	70
Concern	2 (8)	154	77
Uncertainty	1 (4)	67	67
Experts	2 (8)	18	9
Regional and local health departments			
Corrective	4 (3)	119	30
Risk	5 (4)	20	4
Efficacy	81 (60)	711	9
Concern	24 (18)	236	10
Uncertainty	10 (7)	90	9
Experts	12 (9)	35	3
Provincial health authorities			
Corrective	1 (2)	0	0
Risk	12 (23)	38	3
Efficacy	18 (34)	129	7
Concern	10 (19)	220	22
Uncertainty	3 (6)	5	2
Experts	9 (17)	38	4
Medical officers of health			
Corrective	4 (4)	767	192
Risk	24 (22)	2709	113
Efficacy	49 (45)	12,413	253
Concern	13 (12)	3928	302
Uncertainty	4 (4)	1433	358
Experts	15 (14)	1695	113
Provincial health ministers			
Corrective	0 (0)	N/A[c]	N/A
Risk	2 (20)	123	62
Efficacy	3 (30)	175	58
Concern	1 (10)	43	43
Uncertainty	0 (0)	N/A	N/A
Experts	4 (40)	53	13

[a]Risk communication strategies were not mutually exclusive; therefore, a single tweet could contain multiple strategies.

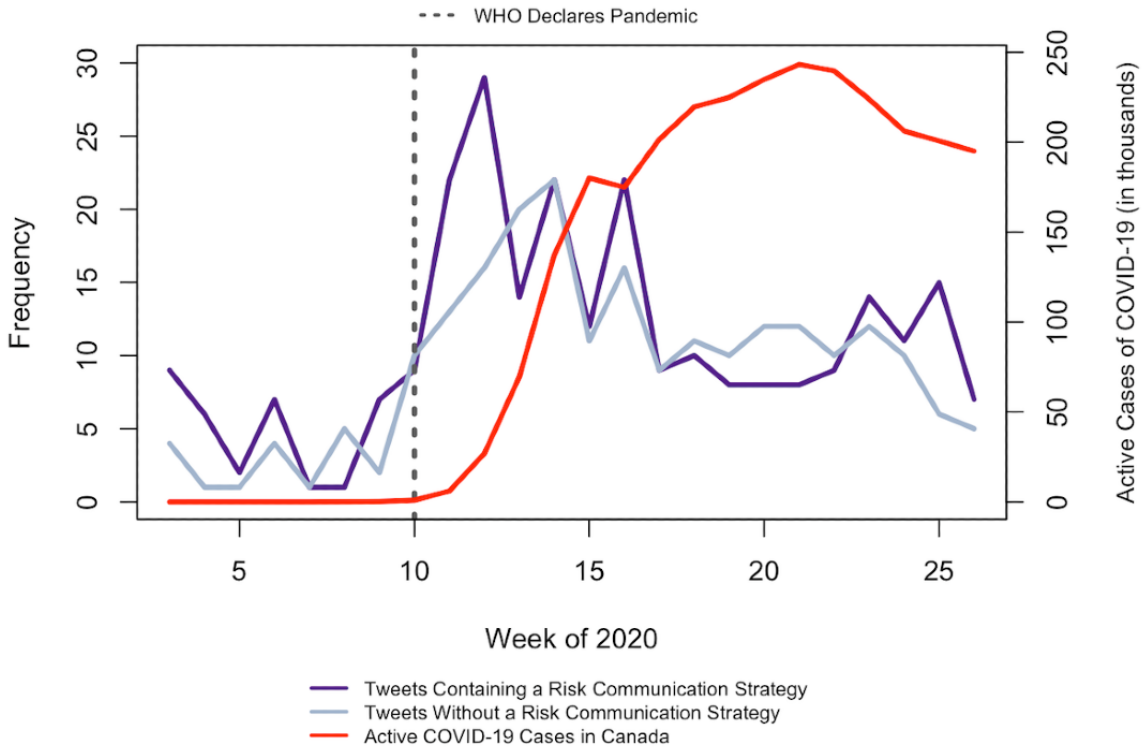
[b]Equals the number of tweets and/or tweet threads authored by Twitter accounts per type between January 1 and June 30, 2020.

[c]N/A: not applicable.

Figure 3 displays the weekly frequency of COVID-19 tweets containing any risk communication strategy and tweets containing no strategy across all account types for the stratified random sample of COVID-19 tweets that were coded during

the content analysis. When examining trends across all account types, the use of risk communication strategies appeared to increase in tweets produced after the WHO's pandemic declaration (week 10) but decreased in the period of time between the first and second peaks of COVID-19 cases in Canada (weeks 15 and 22, respectively). Tweets that did not contain a risk communication strategy were not tweeted as frequently as tweets that did contain a risk communication strategy shortly after the pandemic was declared. The use of any risk communication strategy by all account types appeared to increase again a little after the second peak of COVID-19 cases in Canada after week 22. Tweets using at least one risk communication strategy (n=262) received an average of 61 retweets per tweet, while tweets using no risk communication strategies received an average of 13. This nearly 5-fold difference was statistically significant (P<.001).

Figure 3. Weekly frequency of COVID-19 tweets containing any risk communication strategy, across all account types, January 1, 2020, to June 30, 2020. WHO: World Health Organization.



Discussion

Principal Results

RQ 1 and 2: Tweets and Retweets Over Time by Account Type

Our study of Canadian public health Twitter accounts revealed that tweeting practices and tweet engagement differed between various agencies and decision makers during the COVID-19 pandemic. Of the five account types that we examined, public health agencies and medical officers of health stood out for their tweeting frequency and the high engagement that their tweets received. These two account types had the largest percentage of tweets about COVID-19 relative to all their total number of tweets, as well as the highest daily rate of COVID-19 tweets. This finding is consistent with their roles as the primary agencies and individuals responsible for implementing and communicating a public health response during health crises in Canada. Our findings showed that these two account types also received the highest average engagement (retweets) during the study period, suggesting that Twitter users also recognize these agencies and individuals as the primary information sources to share with their peers during a public health crisis like COVID-19. These findings are positive in that if public health agencies and medical health officers continue to establish a consistent Twitter presence, attract followers, and engage their Twitter audiences, their communications may continue to be shared widely.

We observed that trends in the daily frequency of tweets by account type over time were generally consistent with changes in public concern and engagement over the course of the pandemic [27]. However, it is interesting to note that during key moments in time where the threats of COVID-19 were increasing, the daily rate of COVID-19 tweets did not always correspond to increased retweet counts, and instead trends in tweeting and retweeting varied by account type. For example, when the first case of COVID-19 was confirmed in Canada on January 25, 2020, public health agencies tweeted the most about COVID-19 and received the most retweets for these tweets compared to all other account types; however, after the WHO declared COVID-19 a pandemic on day 70, the daily rates of COVID-19 tweets for all account types began to increase and nearly converged. Medical officers of health received the most retweets per COVID-19 tweet during this period (day 80). These results suggest that Twitter users may have shifted their engagement from tweets authored by public health agencies to

those by medical health officers; however, this trend was less evident by the time Canada's COVID-19 cases peaked.

It is also interesting to note that daily rates of COVID-19 tweets authored by public health agencies seemed to peak around the same time that key moments related to COVID-19 took place in Canada (ie, after Canada's first case, just before the first peak in COVID-19 cases, and shortly after the second peak in COVID-19 cases), which suggests that trends in case counts may have at least partially informed this account type's tweeting practices. This trend was not seen for medical officers of health, whose daily rates of COVID-19 tweets trended downward after the WHO declared COVID-19 a pandemic, despite subsequent peaks in Canadian COVID-19 cases. This is important because increasing COVID-19 case counts partially inform the public about their risk of contracting COVID-19 and thus may lead to increased information seeking. As a result, more public health accounts should respond to increasing case numbers with increased daily tweeting about COVID-19, not less. Unfortunately, information seeking cannot be indicated by engagement metrics such as retweet counts; therefore, our results do not reflect which tweets were most seen or read by Twitter users. However, other research has noted that frequently updated social media feeds are perceived by audiences as more relevant during crises and should be updated enough so as to be noticed by users, regardless of whether they receive engagement [20].

Despite having lower daily COVID-19 tweet rates compared to public health agencies, medical officers of health and provincial health ministers received the highest daily retweets per tweet on day 80, 10 days after the WHO declared COVID-19 a pandemic. This finding was somewhat surprising given that past research has identified health agencies (especially at the federal level) as being able to garner more retweets than other public health accounts during past pandemics, largely due to their recognition and influence over online information sharing [33]. This finding may reflect a Twitter audience shift from governmental institutions to individuals, consistent with some research that has identified a spokesperson effect during public health crises, whereby people seek out a leading voice that is credible and relatable as their primary source of information [43].

One reason that medical officers of health may have been able to maintain a higher average retweet count for the remainder of the study period is that Twitter users may have perceived their content as more medically or scientifically relevant during the pandemic, which other studies have found can lead to more retweets [14,44], likely because expert accounts are perceived as highly authoritative and trustworthy information sources. Perhaps unsurprisingly, Twitter accounts corresponding to provincial health ministers had the lowest percentage of tweets about COVID-19, and the number of retweets that their tweets received decreased significantly after peaking on day 80. Given that provincial health ministers are elected public officials, often with little to no expertise in public health, our findings suggest that these individuals likely relegated responsibilities around COVID-19 communications to other accounts more focused on public health. These results are also consistent with prior research that found that the public may be more likely to distrust information from governments or elected officials and more likely to share information from sources perceived as more trusted or more "expert" (eg, physicians and medical researchers) [45].

RQ3: Use of Engagement Strategies

The frequent use of hashtags and URLs in the majority of COVID-19 tweets that we analyzed is consistent with other studies examining engagement to tweets authored by health agencies [22], and suggests that many Canadian public health agencies and decision makers are aware of and incorporate common Twitter engagement strategies in their tweeting practices. However, the use of these metrics was not always associated with higher engagement, suggesting that there is no single strategy to garner engagement. Hence, each account type may need to tailor their approach. Although user mentions were among the least used engagement metric, and received the lowest average retweets per tweet, the value of these types of tweets extends beyond engagement. In fact, user mentions are viewed as a way to establish dialogue between a tweeter and their audience, build relationships between users, and improve transparency and trust in tweeting institutions [46]. Other studies have also noted that organizations that do not fully utilize Twitter engagement strategies may be missing important opportunities to craft more interactive and engaging communications during a crisis [47].

RQ4: Tweet Message Functions and Use of Risk Communication Strategies

In our content analysis of a stratified random sample of tweets by Canadian public health agencies and decision makers, we found that of the three message functions that we examined, information tweets were most common across all account types, except regional and local public health departments, which used action tweets more frequently. This finding was consistent with other studies that have found tweets conveying information to be the most frequently tweeted by health organizations during other pandemics [26,48]; however, action tweets were most frequently produced in a study by Wahbeh et al [49] of physicians' tweets on COVID-19. Despite information tweets being the most frequently used by the accounts in our sample, action tweets received on average more retweets per tweet for all account types except for public health agencies, which received on average more retweets for information tweets. Other research has also found that action tweets receive the most engagement compared to tweets with other message functions [21]. Our findings suggest that users may seek out and engage with different messages from different account types, relying more on public health agencies for information about COVID-19 and relying on the other accounts for instructions on actions and preventative measures they should be taking. During regular noncrisis periods, community tweets can help a public health agency build an online community and initiate a sense of togetherness [25]; however, the lack of engagement that these tweets received in our study suggest that the public's need for information and direct actions during the COVID-19 pandemic may require public health agencies and decision makers to shift away from community-type tweets during a crisis to meet the needs of their audiences.

Furthermore, the results of our content analysis demonstrated that the risk communication strategies that we examined were not very widely used by any account type, appearing in just over half of the tweets that we analyzed. For example, our study found very few tweets provided corrective information that could be used to tackle misinformation about COVID-19, which is consistent with work by Guidry et al [28] that found only 1% of tweets by health organizations in their sample addressed misinformation on Ebola. These findings suggest that a lack of corrective tweets could represent a missed opportunity for public health agencies and officials to combat misinformation spread during a pandemic. It is also worth noting that risk tweets containing statements that would aid users in making a judgment about their risk of contracting COVID-19 or the harms associated with COVID-19 only appeared in approximately 11% of the tweets in our sample (n=51). In fact, only one category of the risk communication strategies that we examined (efficacy) appeared in at least one-third of the tweets authored by all account types, a frequency that was similar to the percentage of tweets containing efficacy statements about the Zika virus in a study of tweets authored by US public health agencies [33]. On the other hand, tweets that acknowledged concerns about COVID-19 tended to receive among the highest retweets per tweet in our study, which is consistent with risk communication literature identifying concern as an important strategy that aids the public in developing faith in communicators that demonstrate compassion [31].

Despite the overall lack of risk communication strategies employed in the tweets in our sample, our findings demonstrate that including one or more strategies was associated with more engagement on average compared to tweets that did not contain any risk communication strategies (61 retweets per tweet versus 13 retweets per tweet, respectively). We also found that risk communication strategies tended to be used at key moments during the COVID-19 pandemic. The use of risk communication strategies appeared to peak in the weeks just after the WHO declared COVID-19 a pandemic and trended slightly upward after each of the two peaks of COVID-19 cases in Canada. This finding is consistent with other studies that have found that risk communication becomes less prevalent over the course of a crisis [50], since this information is considered most valuable at the beginning of a crisis when uncertainty is high [35]. However, given that our understanding of COVID-19 transmission and health impacts is still developing, a lack of sustained risk communication in tweets by Canadian public health agencies and decision makers could be problematic if it leads to inaccurate perceptions of personal health risks or indifference toward public health measures.

With so much discussion of the pandemic online, supplying information users with high-quality, consistent messaging on the health risks associated with COVID-19 is critical for improving health literacy in the population [51]. Therefore, while the use of these risk communication strategies at key moments could be viewed as promising, more risk communication overall should be undertaken by all public health Twitter accounts to ensure that their audience continuously receives relevant, accurate, and up-to-date information on potential health risks related to COVID-19.

Strengths, Limitations, and Future Work

One methodological advantage that sets our study apart from others is the use of tweet threads, in addition to individual tweets, as the unit of analysis, rather than analyzing each tweet from a thread on its own. Analyzing threads allowed for a more accurate examination of tweeting practices by recreating the message as it would have been viewed originally to a Twitter user. This was a major strength of our content analysis as it allowed for the entire message to be coded and analyzed, rather than a small segment of it. Since threads are commonly used to craft messages that would otherwise be impossible to fit within a tweet's character limit, analyzing them individually would have provided an incomplete picture.

This research has several limitations related, in large part, to its reliance on Twitter data. First, we analyzed the tweets of public health agencies and decision makers in Canada who had tweeted between January 1 and June 30, 2020; however, there are numerous other authoritative health-related Twitter accounts and other official government accounts that tweeted about COVID-19 during this period (eg, physicians, health nongovernmental organizations, etc), which the public may have also engaged with. Moreover, not all members of the public use Twitter; therefore, engagement as measured by counts of retweets does not offer insights into which public health agencies or decision makers the broader Canadian public may consider most important. Additionally, our results may not reflect today or tomorrow's Twitter landscape, and therefore only indicate tweeting practices at a snapshot in time. However, our study can offer a glimpse into trends on information sharing during the first 6 months of the COVID-19 pandemic based on what Canadian public health agencies and decision makers were tweeting and how Twitter users engaged with this content. Our study's findings can be useful to those public health accounts that we included in our analysis as well as to other health organizations or individuals that may be looking for ways to better utilize Twitter to engage with users seeking health information on this platform.

The findings from our study could be improved through additional analysis of content authored by public health accounts on other social media platforms (eg, Instagram, Facebook, etc) or additional refinement of the categories we used for classifying the Twitter accounts and tweet content. For example, future work examining public health communications in Canada could build on our work by contrasting tweeting practices by province or other geographic elements, which could uncover more trends in information sharing and engagement given the region-specific administration of Canada's public health policies. In addition, future work in this area could explore patterns in retweeted tweets and perform a network analysis to examine the Twitter interactions between various public health accounts.

Conclusions

This study analyzed tweets by Canadian public health agencies and decision makers between January 1, 2020, and June 30, 2020, to examine their tweeting practices during the early stages of the pandemic. We also aimed to identify ways that tweets could be improved to effectively communicate risk and maximize engagement on this platform. Using a mixed methods

approach, we conducted Twitter analytics and a qualitative content analysis to characterize the level of engagement that COVID-19 tweets authored by Canadian public health accounts received, as well as the purpose of their tweets and their use of key risk communication strategies. Our research findings suggest that while public health agencies authored more daily COVID-19–related tweets than other account types, engagement tended to be higher for tweets authored by medical health officers, particularly during key moments of the pandemic. Overall, most account types appeared to focus on disseminating information, with the exception of regional and local health departments, which tended to promote more action from users. Our results also point to the need for public health agencies and decision makers to monitor Twitter analytics in order to understand their audience and leverage whatever Twitter engagement strategies help maximize shares of their communications.

During the beginning of any crisis, the use of risk communication strategies by organizations and officials leading the response is critical to help inform the public about an often highly uncertain and rapidly evolving situation, address concerns, and instill trust in those leaders. Our study found that risk communication strategies were not widely used in tweets by any account type, even though these strategies were associated with more engagement. These findings suggest that Canadian public health agencies and decision makers may be missing an important opportunity to engage with information users about the mitigation of health risks related to COVID-19. Finally, our study builds on other literature that has explored differences in engagement to communications authored by individuals and institutions, and suggests that any medical officer of health or other expert individual currently not on Twitter should consider the platform as a means to disseminate information that Twitter users appear to be interested in sharing.

Acknowledgments

This project is supported in part by funding from the Canadian Social Sciences and Humanities Research Council (grant: 435-2020-0257).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flow chart of Tweet exclusion process.
[PNG File , 77 KB - jmir_v23i3e24883_app1.png]

Multimedia Appendix 2
Dataset of manually coded Tweets (n=501) from content analysis.
[XLSX File (Microsoft Excel File), 55 KB - jmir_v23i3e24883_app2.xlsx]

References

1. Coronavirus disease 2019 (COVID-19): Epidemiology update. Government of Canada. 2020. URL: https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html [accessed 2020-10-26]
2. Bronca T. COVID-19: A Canadian timeline. Canadian Healthcare Network. 2020 Apr 8. URL: https://www.canadianhealthcarenetwork.ca/covid-19-a-canadian-timeline [accessed 2020-10-05]
3. French Bourgeois L, Harell A, Stephenson LB. To Follow or Not to Follow: Social Norms and Civic Duty during a Pandemic. Can J Pol Sci 2020 Jun 08;53(2):273-278. [doi: 10.1017/s0008423920000554]
4. Fournier PJ. What Google data says about Canadian vs. U.S. social distancing efforts. Maclean's. 2020 Apr 6. URL: https://www.macleans.ca/society/technology/what-google-data-says-about-canadian-vs-u-s-social-distancing-efforts/ [accessed 2020-10-05]
5. Tam T. Fifteen years post-SARS: Key milestones in Canada's public health emergency response. Can Commun Dis Rep 2018 May 03;44(5):98-101 [FREE Full text] [doi: 10.14745/ccdr.v44i05a01] [Medline: 31007618]
6. Marchildon G. Canada: Health system review. Health Syst Transit 2013;15(1):1-179 [FREE Full text] [Medline: 23628429]
7. Weeks C. How the provinces differ when it comes to messaging on COVID-19. The Globe and Mail. 2020 Mar 17. URL: https://www.theglobeandmail.com/canada/article-how-the-provinces-differ-when-it-comes-to-messaging-on-covid-19/ [accessed 2020-10-05]
8. Eysenbach G. How to Fight an Infodemic: The Four Pillars of Infodemic Management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
9. Chong YY, Cheng HY, Chan HYL, Chien WT, Wong SYS. COVID-19 pandemic, infodemic and the role of eHealth literacy. Int J Nurs Stud 2020 Aug;108:103644 [FREE Full text] [doi: 10.1016/j.ijnurstu.2020.103644] [Medline: 32447127]
10. The Lancet Infectious Diseases. The COVID-19 infodemic. The Lancet Infectious Diseases 2020 Aug;20(8):875 [FREE Full text] [doi: 10.1016/S1473-3099(20)30565-X] [Medline: 32687807]

11. Mohamed T. Coronavirus pushed Twitter to a record 24% growth in daily users last quarter, but the social media giant swung to a loss. Business Insider. 2020 Apr 30. URL: https://markets.businessinsider.com/news/stocks/twitter-q1-earnings-daily-users-rise-24-percent-swings-loss-2020-4-1029149686 [accessed 2020-10-05]
12. Syn SY, Oh S. Why do social network site users share information on Facebook and Twitter? Journal of Information Science 2015 May 19;41(5):553-569. [doi: 10.1177/0165551515585717]
13. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
14. Hagen L, Keller T, Neely S, DePaula N, Robert-Cooperman C. Crisis Communications in the Age of Social Media: A Network Analysis of Zika-Related tweets. Social Science Computer Review 2017 Aug 21;36(5):523-541. [doi: 10.1177/0894439317721985]
15. Liu B, Kim S. How organizations framed the 2009 H1N1 pandemic via social and traditional media: Implications for U.S. health communicators. Public Relations Review 2011 Sep;37(3):233-244. [doi: 10.1016/j.pubrev.2011.03.005]
16. Lachlan KA, Spence PR, Lin X, Najarian K, Del Greco M. Social media and crisis management: CERC, search strategies, and Twitter content. Computers in Human Behavior 2016 Jan;54:647-652. [doi: 10.1016/j.chb.2015.05.027]
17. Garfin DR, Silver RC, Holman EA. The novel coronavirus (COVID-2019) outbreak: Amplification of public health consequences by media exposure. Health Psychol 2020 May;39(5):355-357 [FREE Full text] [doi: 10.1037/hea0000875] [Medline: 32202824]
18. Yoo E, Rand W, Eftekhar M, Rabinovich E. Evaluating information diffusion speed and its determinants in social media networks during humanitarian crises. Journal of Operations Management 2016 Jul 09;45(1):123-133. [doi: 10.1016/j.jom.2016.05.007]
19. Boyd D, Golder S, Lotan G. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. 2010 Presented at: 2010 43rd Hawaii International Conference on System Sciences; Jan 5-8; Honolulu, HI, USA p. 1-10. [doi: 10.1109/hicss.2010.412]
20. Lin X, Spence PR, Sellnow TL, Lachlan KA. Crisis communication, learning and responding: Best practices in social media. Computers in Human Behavior 2016 Dec;65:601-605. [doi: 10.1016/j.chb.2016.05.080]
21. Park H, Reber BH, Chon M. Tweeting as Health Communication: Health Organizations' Use of Twitter for Health Promotion and Public Engagement. J Health Commun 2016;21(2):188-198. [doi: 10.1080/10810730.2015.1058435] [Medline: 26716546]
22. Bhattacharya S, Srinivasan P, Polgreen P. Engagement with health agencies on twitter. PLoS One 2014 Nov 7;9(11):e112235 [FREE Full text] [doi: 10.1371/journal.pone.0112235] [Medline: 25379727]
23. Thackeray R, Neiger BL, Burton SH, Thackeray CR. Analysis of the purpose of state health departments' tweets: information sharing, engagement, and action. J Med Internet Res 2013 Nov 11;15(11):e255 [FREE Full text] [doi: 10.2196/jmir.3002] [Medline: 24217361]
24. Harris JK, Mueller NL, Snider D, Haire-Joshu D. Local health department use of twitter to disseminate diabetes information, United States. Prev Chronic Dis 2013 May 02;10:E70 [FREE Full text] [doi: 10.5888/pcd10.120215] [Medline: 23639765]
25. Neiger BL, Thackeray R, Burton SH, Thackeray CR, Reese JH. Use of twitter among local health departments: an analysis of information sharing, engagement, and action. J Med Internet Res 2013 Aug 19;15(8):e177 [FREE Full text] [doi: 10.2196/jmir.2775] [Medline: 23958635]
26. Wong R, Harris JK, Staub M, Bernhardt JM. Local Health Departments Tweeting About Ebola: Characteristics and Messaging. J Public Health Manag Pract 2017 Sep 2;23(2):e16-e24. [doi: 10.1097/PHH.0000000000000342] [Medline: 26334537]
27. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
28. Guidry JP, Jin Y, Orr CA, Messner M, Meganck S. Ebola on Instagram and Twitter: How health organizations address the health crisis in their social media engagement. Public Relations Review 2017 Sep;43(3):477-486. [doi: 10.1016/j.pubrev.2017.04.009]
29. Odlum M, Yoon S. What can we learn about the Ebola outbreak from tweets? Am J Infect Control 2015 Jun;43(6):563-571 [FREE Full text] [doi: 10.1016/j.ajic.2015.02.023] [Medline: 26042846]
30. Sell TK, Hosangadi D, Trotochaud M. Misinformation and the US Ebola communication crisis: analyzing the veracity and content of social media messages related to a fear-inducing infectious disease outbreak. BMC Public Health 2020 May 07;20(1):550 [FREE Full text] [doi: 10.1186/s12889-020-08697-3] [Medline: 32375715]
31. Seeger M. Best Practices in Crisis Communication: An Expert Panel Process. Journal of Applied Communication Research 2006 Aug;34(3):232-244. [doi: 10.1080/00909880600769944]
32. Sheppard B, Janoske M, Liu B. Understanding Risk Communication Theory: A Guide for Emergency Managers and Communicators. College Park, Maryland: National Consortium for the Study of Terrorism and Responses to Terrorism; 2012 May. URL: https://www.start.umd.edu/sites/default/files/files/publications/UnderstandingRiskCommunicationTheory.pdf [accessed 2021-03-05]
33. Vos SC, Sutton J, Yu Y, Renshaw SL, Olson MK, Gibson CB, et al. Retweeting Risk Communication: The Role of Threat and Efficacy. Risk Anal 2018 Dec 06;38(12):2580-2598. [doi: 10.1111/risa.13140] [Medline: 30080933]

34. Vos SC, Buckner MM. Social Media Messages in an Emerging Health Crisis: Tweeting Bird Flu. J Health Commun 2016;21(3):301-308. [doi: 10.1080/10810730.2015.1064495] [Medline: 26192209]
35. Meadows C, Meadows C, Tang L, Liu W. Unraveling Public Health Crises Across Stages: Understanding Twitter Emotions and Message Types During the California Measles Outbreak. Communication Studies 2019 Mar 24;70(4):453-469. [doi: 10.1080/10510974.2019.1582546]
36. Comrie E, Burns C, Coulson A, Quigley J, Quigley K. Rationalising the use of Twitter by official organisations during risk events: Operationalising the Social Amplification of Risk Framework through causal loop diagrams. European Journal of Operational Research 2019 Jan;272(2):792-801. [doi: 10.1016/j.ejor.2018.07.034]
37. Lwin MO, Lu J, Sheldenkar A, Schulz PJ. Strategic Uses of Facebook in Zika Outbreak Communication: Implications for the Crisis and Emergency Risk Communication Model. Int J Environ Res Public Health 2018 Sep 10;15(9) [FREE Full text] [doi: 10.3390/ijerph15091974] [Medline: 30201929]
38. Structural Profile of Public Health in Canada. National Collaborating Centre for Healthy Public Policy. 2018. URL: http://www.ncchpp.ca/en/structuralprofile.aspx [accessed 2020-10-05]
39. Kearney M, Heiss A, Briatte F. rtweet: Collecting Twitter Data. The R Project for Statistical Computing. 2020. URL: https://CRAN.R-project.org/package=rtweet [accessed 2021-03-08]
40. Lovejoy K, Saxton G. Information, Community, and Action: How Nonprofit Organizations Use Social Media. J Comput-Mediat Commun 2012;17(3):337-353. [doi: 10.1111/j.1083-6101.2012.01576.x]
41. Krippendorff K. Computing Krippendorff's Alpha-Reliability. Penn Libraries. 2011. URL: https://repository.upenn.edu/asc_papers/43 [accessed 2020-10-05]
42. Gamer M, Lemon J, Singh IFP. irr: Various Coefficients of Interrater Reliability and Agreement. The R Project for Statistical Computing. 2019. URL: https://CRAN.R-project.org/package=irr [accessed 2020-10-05]
43. Jin Y, Austin L, Vijaykumar S, Jun H, Nowak G. Communicating about infectious disease threats: Insights from public health information officers. Public Relations Review 2019 Mar;45(1):167-177. [doi: 10.1016/j.pubrev.2018.12.003]
44. Pulido CM, Villarejo-Carballido B, Redondo-Sama G, Gómez A. COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information. International Sociology 2020 Apr 15;35(4):377-392. [doi: 10.1177/0268580920914755]
45. Tang L, Bie B, Park S, Zhi D. Social media and outbreaks of emerging infectious diseases: A systematic review of literature. Am J Infect Control 2018 Sep;46(9):962-972 [FREE Full text] [doi: 10.1016/j.ajic.2018.02.010] [Medline: 29628293]
46. Waters RD, Jamal JY. Tweet, tweet, tweet: A content analysis of nonprofit organizations' Twitter updates. Public Relations Review 2011 Sep;37(3):321-324. [doi: 10.1016/j.pubrev.2011.03.002]
47. Gurman TA, Ellenberger N. Reaching the global community during disasters: findings from a content analysis of the organizational use of Twitter after the 2010 Haiti earthquake. J Health Commun 2015;20(6):687-696. [doi: 10.1080/10810730.2015.1018566] [Medline: 25928401]
48. Hagen L, Neely S, Scharf R, Keller TE. Social Media Use for Crisis and Emergency Risk Communications During the Zika Health Crisis. Digit. Gov.: Res. Pract 2020 Apr 17;1(2):1-21. [doi: 10.1145/3372021]
49. Wahbeh A, Nasralah T, Al-Ramahi M, El-Gayar O. Mining Physicians' Opinions on Social Media to Obtain Insights Into COVID-19: Mixed Methods Analysis. JMIR Public Health Surveill 2020 Jun 18;6(2):e19276 [FREE Full text] [doi: 10.2196/19276] [Medline: 32421686]
50. Spence PR, Lachlan KA, Lin X, del Greco M. Variability in Twitter Content Across the Stages of a Natural Disaster: Implications for Crisis Communication. Communication Quarterly 2015 Apr 06;63(2):171-186. [doi: 10.1080/01463373.2015.1012219]
51. Rudd R, Baur C. Health literacy and early insights during a pandemic. Journal of Communication in Healthcare 2020 May 13;13(1):13-16. [doi: 10.1080/17538068.2020.1760622]

Abbreviations

ANOVA: analysis of variance
PHAC: Public Health Agency of Canada
RQ: research question
SARS: severe acute respiratory syndrome
WHO: World Health Organization

Edited by G Eysenbach, C Basch; submitted 08.10.20; peer-reviewed by D Wawrzuta, Z Shah; comments to author 23.10.20; revised version received 30.10.20; accepted 17.02.21; published 11.03.21.

Please cite as:
Slavik CE, Buttle C, Sturrock SL, Darlington JC, Yiannakoulias N
Examining Tweet Content and Engagement of Canadian Public Health Agencies and Decision Makers During COVID-19: Mixed Methods Analysis
J Med Internet Res 2021;23(3):e24883
URL: https://www.jmir.org/2021/3/e24883
doi:10.2196/24883
PMID:33651705

Viewpoint

Clinical Trial Data Sharing for COVID-19–Related Research

Louis Dron[1], MSc; Alison Dillman[2], MPH; Michael J Zoratti[1], MSc; Jonas Haggstrom[3], PhD; Edward J Mills[1], FRCP Edin; Jay J H Park[4], MSc

[1]Cytel Canada Inc., Vancouver, BC, Canada

[2]School of Public Health, Faculty of Medicine, Imperial College London, London, United Kingdom

[3]International COVID-19 Data Alliance, London, United Kingdom

[4]Department of Experimental Medicine, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Jay J H Park, MSc
Department of Experimental Medicine, University of British Columbia
802-777 West Broadway
Vancouver, BC, V5Z 1J5
Canada
Phone: 1 (604) 294 3823
Email: jayhpark1@alumni.ubc.ca

Abstract

This paper aims to provide a perspective on data sharing practices in the context of the COVID-19 pandemic. The scientific community has made several important inroads in the fight against COVID-19, and there are over 2500 clinical trials registered globally. Within the context of the rapidly changing pandemic, we are seeing a large number of trials conducted without results being made available. It is likely that a plethora of trials have stopped early, not for statistical reasons but due to lack of feasibility. Trials stopped early for feasibility are, by definition, statistically underpowered and thereby prone to inconclusive findings. Statistical power is not necessarily linear with the total sample size, and even small reductions in patient numbers or events can have a substantial impact on the research outcomes. Given the profusion of clinical trials investigating identical or similar treatments across different geographical and clinical contexts, one must also consider that the likelihood of a substantial number of false-positive and false-negative trials, emerging with the increasing overall number of trials, adds to public perceptions of uncertainty. This issue is complicated further by the evolving nature of the pandemic, wherein baseline assumptions on control group risk factors used to develop sample size calculations are far more challenging than those in the case of well-documented diseases. The standard answer to these challenges during nonpandemic settings is to assess each trial for statistical power and risk-of-bias and then pool the reported aggregated results using meta-analytic approaches. This solution simply will not suffice for COVID-19. Even with random-effects meta-analysis models, it will be difficult to adjust for the heterogeneity of different trials with aggregated reported data alone, especially given the absence of common data standards and outcome measures. To date, several groups have proposed structures and partnerships for data sharing. As COVID-19 has forced reconsideration of policies, processes, and interests, this is the time to advance scientific cooperation and shift the clinical research enterprise toward a data-sharing culture to maximize our response in the service of public health.

(J Med Internet Res 2021;23(3):e26718) doi:10.2196/26718

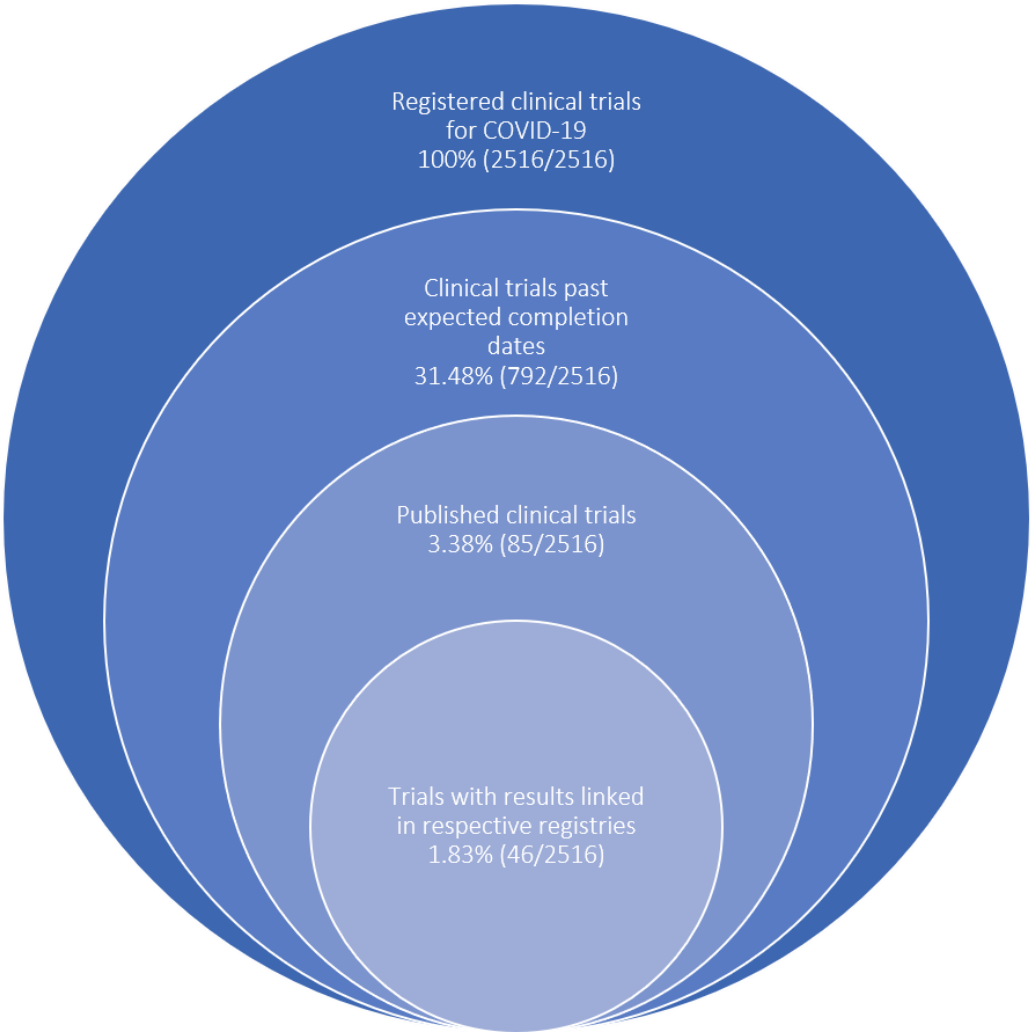
KEYWORDS

COVID-19; data-sharing; clinical trials; data; research; privacy; security; registry; feasibility; challenge; recruitment; error; bias; assessment; interoperability; dataset; intervention; cooperation

The scientific community has made several important inroads in the fight against COVID-19. The pandemic has mobilized the global research community at an unparalleled scale [1-3]. From the start of the COVID-19 pandemic to date, 2516 clinical trials have been registered globally [4]. Most are within hospitalized patient contexts, with other trials focusing on outpatient treatment or prophylaxis, whether through vaccination or pre- or posttreatment prophylaxis. Of the 2516 registered clinical trials, records indicate that 1278 (50.79%) trials are still actively enrolling patients, 26 (1.03%) have suspended recruitment, 43 (1.70%) have terminated, and 67 (2.66%) have withdrawn [4]. However, it is important to note that the status of 28.22% (710/2516) of these trials have not been updated in their respective registries since they were first posted, whereas only 1.83% (46/2516) of the trials that are past their expected completion dates have reported results linked to their respective

registries [4]. According to a living systematic review on randomized clinical trials for COVID-19 published in *The BMJ*, only 85 trials have been published as of October 21, 2020, despite the large number of trials that have been registered and reported as complete [5]. Of these 85 published trials, 54 (64%) trials reported information on planned sample size, and 25 (46%) did not meet their recruitment targets [5]. In fact, they recruited approximately half of their planned recruitment (median 52.3%; IQR 31.7%-80.6%) [5]. A summary of these findings is provided in Figure 1.

Figure 1. Summary of registered clinical trial status for COVID-19–related research.



During this time, we have seen a large number of trials conducted without results being made available. It is likely that a plethora of COVID-19 trials have stopped early, not for statistical reasons but due to lack of feasibility [4]. Reasons for studies becoming nonfeasible are extensive, ranging from unwillingness to participate due to quarantine, challenges in telemedicine solutions for trials [6], and emergency changes to staff resourcing [7]. Furthermore, there are likely feasibility challenges in the recruitment dependent on the patient context. Trials range from patients in intensive care to healthy volunteers in vaccine and treatment prophylaxis trials. As such, contributions of recruitment competition and patient hesitancy to the lack of feasibility are heavily treatment-context driven. Accordingly, many clinical trials during the COVID-19 pandemic have faced a multitude of challenges related to consenting and recruiting new participants given the proliferation of trials that are competing for recruiting eligible participants into their own respective trials [8,9].

Trials stopped early for feasibility are, by definition, statistically underpowered and thereby prone to inconclusive findings. Statistical power is not necessarily linear with the total sample size, and even small reductions in patient numbers or events can have a substantial impact on the research outcomes. Given the profusion of clinical trials investigating identical or similar treatments across different geographical and clinical contexts, one must also consider that the likelihood of the substantial numbers of false-positive and false-negative trials, emerging with the increasing overall number of trials, adds to public perceptions of uncertainty. Complicating this issue is the evolving nature of the pandemic, where baseline assumptions on control group risk factors used to develop sample size calculations are far more challenging than those in the case of well-documented diseases.

The standard answer to address these challenges during nonpandemic settings is to rigorously assess each trial for statistical power and risk-of-bias and then pool the reported aggregated results using meta-analytic statistical approaches.

This solution simply will not suffice for COVID-19. Even with random-effects meta-analysis models, it will be difficult to adjust for heterogeneity of different trials with aggregated reported data alone, especially given the absence of common data standards and outcome measures. Common data standards are a key feature of system interoperability and facilitate synthesis methodologies to be rapidly scaled, such as meta-analyses. To date, several groups have proposed structures and partnerships for data sharing in the context of COVID-19, some of which are integrated with prespecified statistical analysis methodologies [10,11]. Given the substantial under-recruitment reported to date, vast numbers of trials will be underpowered. This lack of power may be due to design challenges, or a consequence of termination prior to reaching the recruitment target. As such, integration of different trial datasets for individual participant-level data (IPD) meta-analyses may be the only solution in determining what works and is safe for COVID-19. In an IPD meta-analysis, rather than measuring aggregate study-level outputs, data can be taken from either all or a proportion of participants within individual studies. In doing so, more nuanced comparisons between patient groups is possible. For example, participants across two trials may have, on average, significant differences in demographics to one another, yet substantial proportions of patients across both trials may have sufficient similarity for a valid analysis. Meta-analyses integrating IPD have a number of potential methodological advantages, particularly when subpopulations of interest demonstrate promising treatment signals. In particular, IPD meta-analyses allow for more effective subgroup analyses and better statistical power for detecting treatment interaction effects [12] in cases wherein differences between populations are marked. These methods are endorsed by the Cochrane collaboration—a useful tool, especially when treatment effects are influenced by the follow-up duration. As COVID-19 research evolves to longer-term outcomes (sometimes referred to as "long-COVID") [13], these analytical advantages are likely to develop further. The key to the process of evidence synthesis is the appropriate selection of trials with comparable patient populations and design features, such as outcome definitions [14]. In the absence of unified data structures and data sharing agreements, this process may either be time consuming or entirely nonfeasible, depending on data heterogeneity.

To serve the public who are waiting for the medical research community to efficiently make medical discoveries, the COVID-19 pandemic has (arguably) mandated sharing IPD into a public health obligation. Although the International Committee of Journal Editors has previously discussed the importance of data sharing of clinical trials, this discussion has largely been limited to published clinical trials [15]. The data sharing mandate for the COVID-19 pandemic should be extended to all clinical trials, including those trials that will not be published because they ended early for feasibility reasons. Other informal data sharing platforms for ongoing trials are available, such as clinical trial registries, which can provide information on outcome measures and broad design features [4].

Sharing of IPD has historically proven to be challenging, as investigators and sponsors have held tight to their data for academic, regulatory, and commercial reasons. However, the health and economic consequences of the pandemic thus far signal a need to mandate data sharing, expedite systems to apportion credit for data sharing, and preserve commercial interests. The need to share and collaborate openly supersedes our personal career or organizational goals. This sentiment has been shared among the research community, and many organizations (such as the Wellcome trust [16]) have quickly identified the need to share data more rapidly than has historically been the case.

The International COVID-19 Data Alliance (ICODA) provides an example of recent improvements in data sharing within the context of a global pandemic [17,18]. Convened by Health Data Research UK, ICODA is an international health data-led research response that seeks to provide a platform to enable researchers to access global data to derive insights about COVID-19 to advance the development of therapeutics [17,18]. The organization recognizes the urgent need to enable access to data that can be linked with other data in a safe and secure way.

As the processes for addressing personal privacy, data security, and data standardization have become sufficiently more sophisticated in recent years, barriers previously considered to be insurmountable have been minimized [10,19]. Investigators that have launched clinical trials can utilize existing global clinical research data sharing platforms such as Vivli [19], TransCelerate Biopharma [20], and ICODA [17,18], to collectively and securely curate and analyze their findings. Taken together, data from different trials can answer meaningful public health questions while avoiding the risk of becoming inconclusive in isolation. Investigators are keen for data; as represented by Vivli [11], as of December 2020, over 200 requests for trial-level data have been made in 2020, although no publications utilizing COVID-19 data are available from this group at present. Challenges in execution of these methods and strategies are multifaceted, involving researcher awareness of resource availability, technical capacity for analysis, and access agreements from data providers. To this end, we applaud the efforts of groups mentioned above in reaching out to researchers for proposals and taking strides toward simplifying the often challenging process of providing patient-level data. Successful analysis and subsequent publications utilizing these methods may provide an informative case study to promote further researcher contribution.

Particularly in the context of a pandemic, researchers, policymakers, and the general public are finding challenges in navigating the multitudes of data available daily. In tandem, high-profile instances of retractions owing to poor data screening [21] have led many to reach "data fatigue" [22]. Here, data synthesis exercises that utilize the aforementioned statistical efficiencies of patient-level data provide an avenue through which data fatigue may be minimized and succinct summaries that may otherwise be unachievable, as well as improve awareness of therapeutic trends in COVID-19.

We hope that the COVID-19 pandemic is a historic turning point of a sharing culture in the medical research community. The need for rapid and robust clinical research for the discovery of effective and safe therapeutics and vaccines has never been

higher. Strengthening our public health response to COVID-19 will require larger collated patient-level datasets to facilitate the scientific precision required for answers on COVID-19 medical interventions. As COVID-19 has forced reconsideration of policies, processes, and interests, this is the time to advance scientific cooperation and shift the clinical research enterprise toward a data-sharing culture that can maximize our response in the service of public health.

Conflicts of Interest

None declared.

References

1. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Apr;76:71-76 [FREE Full text] [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]
2. Fauci A, Lane H, Redfield R. Covid-19 - navigating the uncharted. N Engl J Med 2020 Mar 26;382(13):1268-1269 [FREE Full text] [doi: 10.1056/NEJMe2002387] [Medline: 32109011]
3. Velavan TP, Meyer CG. The COVID-19 epidemic. Trop Med Int Health 2020 Mar;25(3):278-280 [FREE Full text] [doi: 10.1111/tmi.13383] [Medline: 32052514]
4. Thorlund K, Dron L, Park J, Hsu G, Forrest JI, Mills EJ. A real-time dashboard of clinical trials for COVID-19. The Lancet Digital Health 2020 Jun;2(6):e286-e287. [doi: 10.1016/s2589-7500(20)30086-8]
5. Siemieniuk RA, Bartoszko JJ, Ge L, Zeraatkar D, Izcovich A, Kum E, et al. Drug treatments for covid-19: living systematic review and network meta-analysis. BMJ 2020 Jul 30;370:m2980 [FREE Full text] [doi: 10.1136/bmj.m2980] [Medline: 32732190]
6. Liu N, Huang R, Baldacchino T, Sud A, Sud K, Khadra M, et al. Telehealth for noncritical patients with Chronic diseases during the COVID-19 pandemic. J Med Internet Res 2020 Aug 07;22(8):e19493 [FREE Full text] [doi: 10.2196/19493] [Medline: 32721925]
7. Mascha E, Schober P, Schefold J, Stueber F, Luedi MM. Staffing with disease-based epidemiologic indices may reduce shortage of intensive care unit staff during the COVID-19 pandemic. Anesth Analg 2020 Jul;131(1):24-30 [FREE Full text] [doi: 10.1213/ANE.0000000000004849] [Medline: 32343514]
8. Mitchell EJ, Ahmed K, Breeman S, Cotton S, Constable L, Ferry G, et al. It is unprecedented: trial management during the COVID-19 pandemic and beyond. Trials 2020 Sep 11;21(1):784 [FREE Full text] [doi: 10.1186/s13063-020-04711-6] [Medline: 32917258]
9. Cunniffe NG, Gunter SJ, Brown M, Burge SW, Coyle C, De Soyza A, et al. How achievable are COVID-19 clinical trial recruitment targets? A UK observational cohort study and trials registry analysis. BMJ Open 2020 Oct 05;10(10):e044566 [FREE Full text] [doi: 10.1136/bmjopen-2020-044566] [Medline: 33020111]
10. Petkova E, Antman EM, Troxel AB. Pooling data from individual clinical trials in the COVID-19 era. JAMA 2020 Aug 11;324(6):543-545. [doi: 10.1001/jama.2020.13042] [Medline: 32717043]
11. Cosgriff CV, Ebner DK, Celi LA. Data sharing in the era of COVID-19. The Lancet Digital Health 2020 May;2(5):e224. [doi: 10.1016/s2589-7500(20)30082-0]
12. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ 2010 Feb 05;340:c221. [doi: 10.1136/bmj.c221] [Medline: 20139215]
13. Gorna R, MacDermott N, Rayner C, O'Hara M, Evans S, Agyen L, et al. Long COVID guidelines need to reflect lived experience. The Lancet 2021 Feb;397(10273):455-457. [doi: 10.1016/s0140-6736(20)32705-7]
14. Tierney JF, Fisher DJ, Burdett S, Stewart LA, Parmar MKB. Comparison of aggregate and individual participant data approaches to meta-analysis of randomised trials: An observational study. PLoS Med 2020 Jan;17(1):e1003019 [FREE Full text] [doi: 10.1371/journal.pmed.1003019] [Medline: 32004320]
15. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. JAMA 2017 Jun 27;317(24):2491-2492. [doi: 10.1001/jama.2017.6514] [Medline: 28586895]
16. Carr D. Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak. Wellcome - Press Release. 2020 Jan 31. URL: https://wellcome.ac.uk/coronavirus-covid-19/open-data [accessed 2021-03-10]
17. International COVID-19 Data Alliance (ICODA). URL: https://icoda-research.org/ [accessed 2020-12-22]
18. International COVID-19 Data Alliance (ICODA). Health Data Research UK. URL: https://www.hdruk.ac.uk/covid-19/international-covid-19-data-alliance/ [accessed 2020-12-22]
19. Bierer BE, Li R, Barnes M, Sim I. A global, neutral platform for sharing trial data. N Engl J Med 2016 Jun 23;374(25):2411-2413. [doi: 10.1056/nejmp1605348]
20. COVID-19 - TransCelerate. TansCelerate Biopharma Inc. URL: https://transceleratebiopharmainc.com/covid-19/ [accessed 2021-03-10]

21. Transparency during global health emergencies. The Lancet Digital Health 2020 Sep;2(9):e441. [doi: 10.1016/s2589-7500(20)30198-9]
22. Pandemic fatigue - reinvigorating the public to prevent COVID-19. WHO Regional Office for Europe. 2020 Oct 05. URL: https://www.euro.who.int/en/media-centre/events/events/2020/10/pandemic-fatigue-reinvigorating-the-public-to-prevent-covid-19 [accessed 2021-03-10]

Abbreviations

ICODA: International COVID-19 Data Alliance
IPD: individual participant-level data

Edited by G Eysenbach; submitted 22.12.20; peer-reviewed by N Zariffa, M Fazeli; comments to author 04.01.21; revised version received 08.01.21; accepted 05.03.21; published 12.03.21.

Please cite as:
Dron L, Dillman A, Zoratti MJ, Haggstrom J, Mills EJ, Park JJH
Clinical Trial Data Sharing for COVID-19–Related Research
J Med Internet Res 2021;23(3):e26718
URL: https://www.jmir.org/2021/3/e26718
doi:10.2196/26718
PMID:33684053

Original Paper

A Self-Assessment Web-Based App to Assess Trends of the COVID-19 Pandemic in France: Observational Study

Fabrice Denis[1*], MD, PhD; Arnaud Fontanet[2*], MD, PhD; Yann-Mael Le Douarin[3*], MD; Florian Le Goff[4*], MSc; Stephan Jeanneau[5*], MSc; François-Xavier Lescure[6], MD, PhD

[1]Institut Inter-régional de Cancérologie Jean Bernard, Le Mans, France
[2]Emerging Diseases Epidemiology Unit, Institut Pasteur, Paris, France
[3]French Society of Digital Health, Paris, France
[4]Kelindi, Lille, France
[5]Adobis Group, Grenoble, France
[6]AP-HP, Infectious and Tropical Diseases Department, Bichat-Claude Bernard University Hospital, Paris, France
[*]these authors contributed equally

Corresponding Author:
Fabrice Denis, MD, PhD
Institut Inter-régional de Cancérologie Jean Bernard
9 rue Beauverger
Le Mans, 72000
France
Phone: 33 684190480
Email: f.denis@cjb72.org

Abstract

Background: We developed a self-assessment and participatory web-based triage app to assess the trends of the COVID-19 pandemic in France in March 2020.

Objective: We compared daily large-scale RT–PCR test results to monitor recent reports of anosmia through a web-based app to assess the dynamics of emergency department visits, hospitalizations, and intensive care unit (ICU) admissions among individuals with COVID-19 in France.

Methods: Between March 21 and November 18, 2020, users of the maladiecoronavirus.fr self-triage app were asked questions about COVID-19 symptoms. Data on daily hospitalizations, large-scale positive results on RT–PCR tests, emergency department visits, and ICU admission of individuals with COVID-19 were compared to data on daily reports of anosmia on the app.

Results: As of November 18, 2020, recent anosmia was reported 575,214 times from among approximately 13,000,000 responses. Daily anosmia reports during peak engagement with the app on September 16, 2020, were spatially correlated with the peak in daily COVID-19–related hospitalizations in November 2020 (Spearman rank correlation coefficient [ρ]=0.77; P<.001). This peak in daily anosmia reports was observed primarily among young adults (age range 18-40 years), being observed 49 days before the peak of hospitalizations that corresponded to the first wave of infections among the young population, followed by a peak in hospitalizations among older individuals (aged ≥50 years) in November 2020. The reduction in the daily reports of anosmia associated with the peaks in the number of cases preceded the reduction in daily hospitalizations by 10 and 9 days during the first and the second waves of infection, respectively, although the reduction in the positivity rates on RT–PCR tests preceded the reduction in daily hospitalizations by only 2 days during the second wave of infections.

Conclusions: Data on daily reports of anosmia collected through a nationwide, web-based self-assessment app can be a relevant tool to anticipate surges in outbreaks, hospitalizations, and ICU admission during the COVID-19 pandemic.

Trial Registration: ClinicalTrials.gov NCT04331171; https://clinicaltrials.gov/ct2/show/NCT04331171

(J Med Internet Res 2021;23(3):e26182) doi:10.2196/26182

KEYWORDS

app; big data; COVID-19; diagnosis; diagnostic test; digital health; France; mobile phone; observational; participatory app; self-assessment; surveillance; trend; web-based app

Introduction

Apps involving patient-reported outcomes have been shown to improve outcomes including survival benefit [1-3]. We developed and launched a self-assessment and participatory web-based surveillance app for COVID-19, called "maladiecoronavirus.fr," during the growth phase of the COVID-19 pandemic in March 2020 in France. This self-triage tool was aimed at directing symptomatic patients with COVID-19 to emergency care or to general practitioners after the analysis of symptoms and comorbidities. We previously reported that data from this web-based app could be a relevant tool to reduce the burden on emergency call centers [4]. Interestingly, this web-based app was useful in monitoring COVID-19 spread during the initial wave of infections in March 2020 in France, with spatial correlations among the number of hospitalizations, users reporting fever and cough, and users reporting anosmia [5]. The ability to detect an early rise and decline in COVID-19 incidence would also be extremely useful in anticipating a patient's course of hospitalization to avoid overcrowding at emergency departments and saturation of the intensive care unit (ICU) and to anticipate the availability of hospital beds dedicated to patients with COVID-19.

In France, nationwide RT–PCR test results are available on a daily basis and are used to estimate the effective reproduction number. However, variability in testing indications and access to tests have led to speculations regarding the practicality and validity of large-scale RT–PCR test results and concomitantly the monitoring of trends of the pandemic. Our web-based app may be a useful alternative. In this study, we compared the results of large-scale RT–PCR tests on a daily basis to daily reports of anosmia from among the users of the app to predict the dynamics of emergency department visits, hospitalizations, and ICU admission of individuals testing positive for COVID-19 during the outbreak in France.

Methods

Users of maladiecoronavirus.fr in France were recruited through nationwide media campaigns, including social media, radio, and magazine campaigns, during March 17-29, 2020. Participants were recruited through the maladiecoronavirus.fr website as previously described [5]. Data including sociodemographic characteristics, zip code, and comorbidities were anonymously obtained. The participants were asked about the following symptoms potentially associated with COVID-19: fever (body temperature of >37.7°C), unusual cough, shortness of breath, sore throat, muscle aches, diarrhea, anorexia, and asthenia. Anosmia was included in this list of potential symptoms on March 21, 2020. After recording the symptoms of the participants, a notification was sent to them recommending them to either stay at home and use the website again in case of evolving symptomatology (self-monitoring) or to contact a general practitioner or an emergency number if they reported experiencing dyspnea or severe anorexia. Questionnaires responses were excluded from the analysis if they did not include a zip code or if the duration of completion was inconsistent (<30 seconds). The study was approved by the French Health Data Hub, which reviews the ethical conduct of research with human subjects and the confidentiality and safety of their data. The web-based app was not considered a medical device by regulatory authorities because no tracking was performed, and the data were anonymized. The app did not monitor the adherence of the participants to the self-triage recommendations and did not inquire participants about their test results. Participants did not need to create an account or log into the app to access it. The app did not identify duplicate responses and did not make follow-up inquiries with the participants.

Data on daily large-scale positive outcomes on diagnostic RT–PCR tests, emergency department visits, hospitalizations, and ICU admissions among individuals with COVID-19 were obtained from Santé publique France, the Oscour network, and the French Ministry of Health. Big data were analyzed by Data Chain (Adobis Group). We did not ask app users for daily reports, but we assessed daily overall reports of anosmia among the users. We compared these daily reports of anosmia on the app, daily positive RT–PCR tests results, daily emergency department visits, daily conventional hospitalization, and daily ICU admissions among individuals with COVID-19 in order to determine which source of data best predicts the peaks and declines in hospitalization and ICU admission.

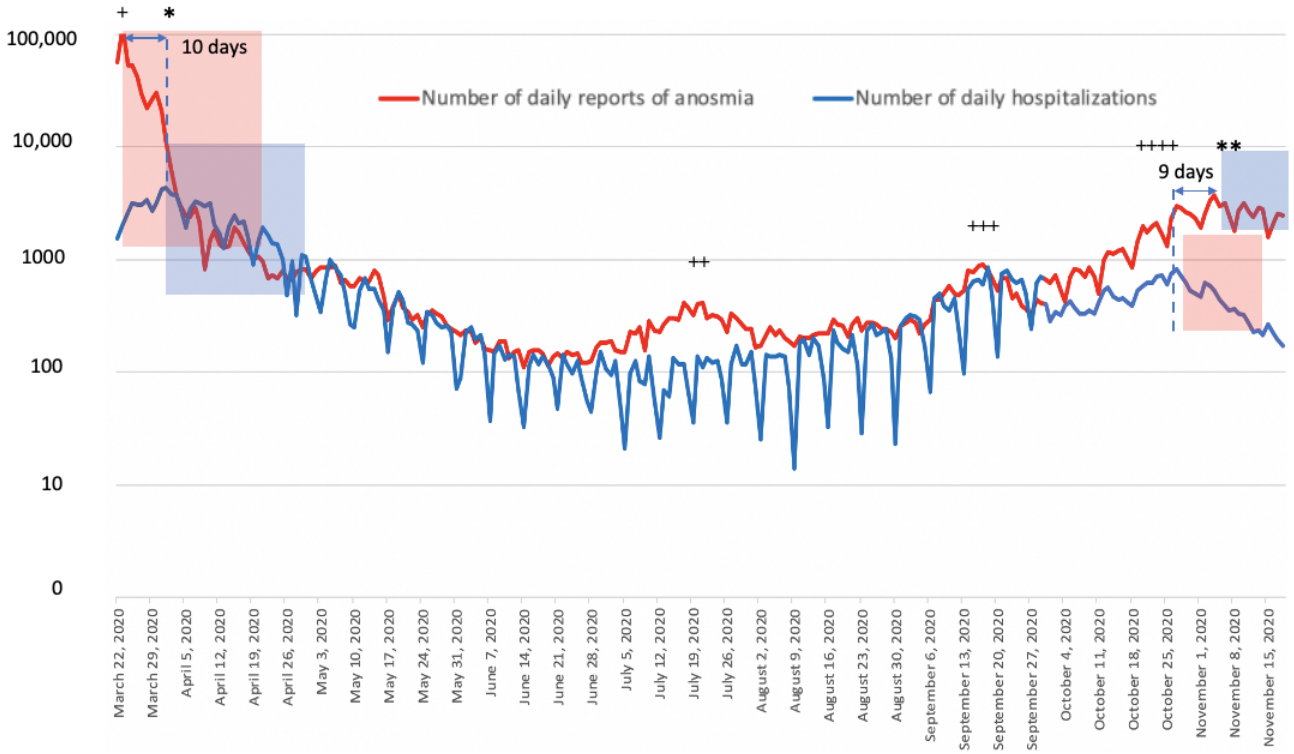
Age stratification was performed to assess the predictability of subsequent hospitalization. Spearman rank correlation analysis was performed for statistical analysis.

Results

Between March 17 and November 18, 2020, a total of 13,000,343 completed questionnaires were included, of which 7,507,332 were excluded owing to the unavailability of the zip code or an inconsistent completion duration, and recent anosmia was reported 575,214 times. The number of assessed questionnaires represents the number of assessments and not the number of individuals.

Data on daily reports of anosmia on the website and daily hospitalizations during the outbreak were well correlated (ρ=0.75; P<.001) (Figure 1). During the first wave of infections in March 2020, the peak in daily reports of anosmia (113,234 connections) was reached on March 22, 2020, and that of daily hospitalizations (4281 patients) was reached 10 days later on April 1, 2020.

Figure 1. Comparisons between the daily number of hospitalizations and the daily reports of anosmia on the web-based self-triage app maladiecoronavirus.fr during the first and second peaks of the pandemic in France between March 21, 2020, and November 18, 2020. The first peak of hospitalizations occurred on January 4, 2020, and the second one occurred on April 11, 2020. In total, 4 peaks of daily reports of anosmia were recorded from the web-based app: March 23, 2020 (+), July 22, 2020 (++), September 16, 2020 (+++), and October 26, 2020 (++++). *Peak of hospitalizations during the first wave of the COVID-19 pandemic. **Peak of hospitalizations during the second wave of the COVID-19 pandemic. The graph shows a reduction in the number of connections 10 days before the reduction in daily hospitalizations during the first wave of the pandemic and 9 days before the reduction in daily hospitalizations during the second wave (semilog scale). Red transparent windows show the reduction in daily reports of anosmia, and blue transparent windows show the reduction in the number of daily hospitalizations during the first and second waves of the pandemic.



Before and during the second wave of infections towards the end of October 2020, a total of 3 early peaks of daily reports of anosmia on the website occurred on July 21, 2020 (n=417); September 16, 2020 (n=905); and October 26, 2020 (n=805); and the peak in daily hospitalizations was reached on November 4, 2020 (n=3681), having occurred 49 days after the highest peak in the daily reports of anosmia. Daily hospitalizations decreased after November 4, 2020; this occurred 9 days after the last peak in daily reports of anosmia. Data on the median age of users of the web-based app are shown in Figure 2. The median age of the participants was 40 years (IQR 27-56) during the first peak of connections (first wave of the pandemic in March 2020); 32 years during the second peak of connections on July 22, 2020; 30 years during the third peak of connections on September 16, 2020; and 37 years during the fourth peak of connections on October 26, 2020 (second wave of the pandemic). Participants aged ≥45 years were more numerous during the last peak of connections in October 2020 than in third peak in September 2020 (n=190/2088, 9.1% vs n=402/1668, 24.1%, respectively) (Figures 2A and 2B).

The number of ICU admissions peaked (n=450) on November 4, 2020, 9 days after the last peak of daily reports of anosmia. Daily numbers of emergency department visits peaked (n=787) on November 2, 2020, a total of 7 days after the last peak of daily reports of anosmia. The number of positive outcomes on daily RT–PCR tests peaked (n=69,564) on November 2, 2020, a total of 7 days after the last peak of daily reports of anosmia and only 2 days before the peak of hospitalization occurred (Figures 3A, 3B, 3C, and 3D).

A large-scale curfew, followed by a general lockdown, was initiated on October 23, 2020. Contamination in RT–PCR tests reduced on November 2, 2020, a total of 15 days after lockdown enforcement, and daily reports of anosmia decreased from October 27, 2020, a total of 4 days after lockdown enforcement.

Large-scale positive outcomes on daily RT–PCR tests and daily reports of anosmia during the third peak of connections on September 16, 2020, were spatially correlated at the county level with the peak in daily COVID-19–related hospitalizations in November 2020 (ρ=0.77 for both; P<.001) (Figure 4).

Figure 2. Age of users reporting anosmia on the web-based self-triage app maladiecoronavirus.fr during the COVID-19 outbreak in France. (2A) Median age of the users of the app. *Peak of hospitalizations during the first wave of the pandemic. **Peak of hospitalizations during the second wave of the pandemic. Four peaks of daily reports of anosmia were recorded from the web-based app: the first peak of reports of anosmia (+; median age=40 years), second peak of reports of anosmia (++; median age=32 years), third peak of reports of anosmia (+++; median age=30 years), and fourth peak of reports of anosmia (++++; median age=37 years). (2B) Histograms of the number of reports of anosmia based on the age of the users during third and fourth peaks of connections. Users of aged >45 years were more numerous in the October 2020 outbreak than in the September 2020 outbreak peak of during the second wave of hospitalizations (n=190/2088, 9.1% vs n=402/1668, 24.1% respectively).

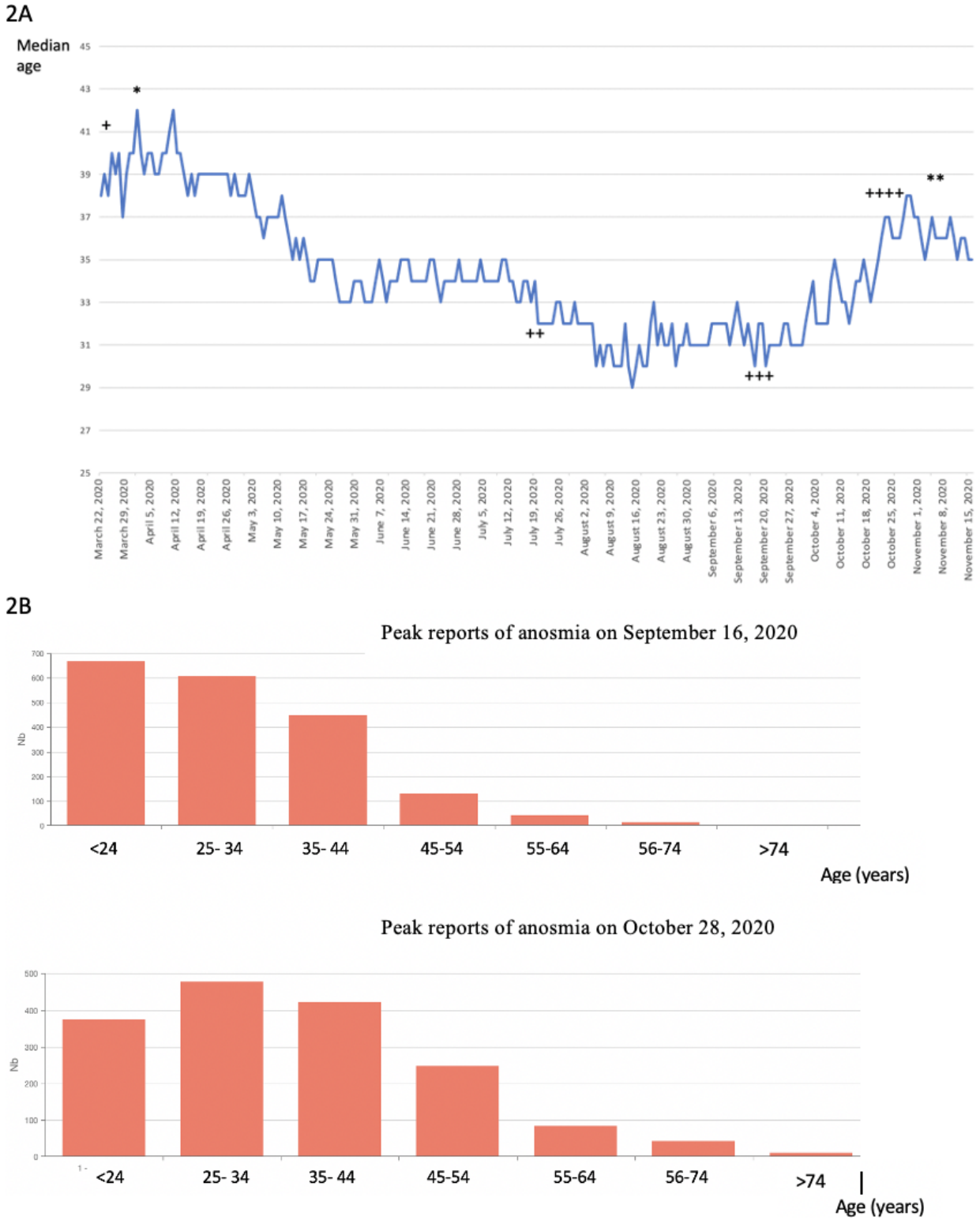


Figure 3. Trends of the COVID-19 pandemic during the second wave of infections in France from August 10 to November 18, 2020: (3A) comparison between the number of daily reports of anosmia and positive outcomes on daily RT–PCR tests; (3B) comparison between the number of daily reports of anosmia and the number of emergency department visits; (3C) comparison between the number of daily reports of anosmia and the number of daily hospitalizations; and (3D) comparison between the number of daily reports of anosmia and the number of daily admissions to the intensive care unit.

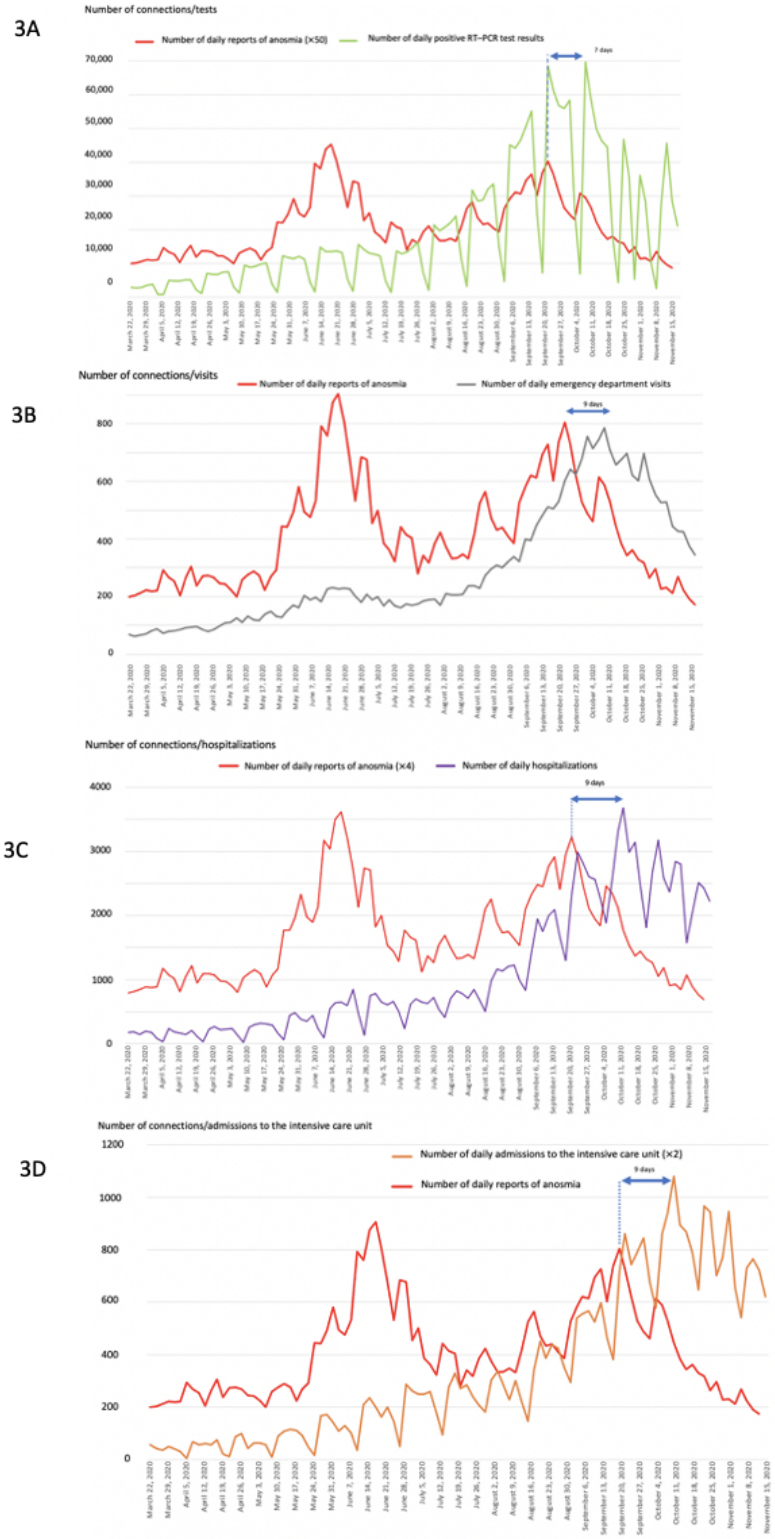
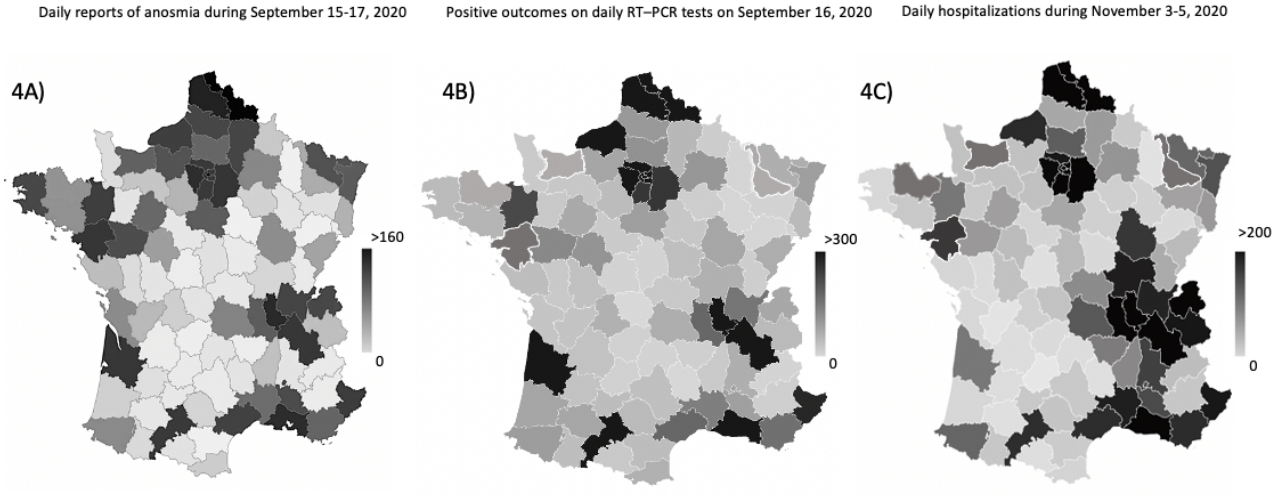


Figure 4. Maps displaying the correlation between daily reports of anosmia with COVID-19–related hospitalizations during the second wave of the pandemic in France. The cumulative number of (4A) daily reports of anosmia during September 15-17, 2020; (4B) positive outcomes on large-scale RT–PCR tests on September 16, 2020; and (4C) daily hospitalizations of individuals with COVID-19 in France during November 3-5, 2020.



Discussion

Our results suggest that the peak of daily reports of anosmia determined from our web-based self-triage app based on symptoms reported by individuals suspected with COVID-19 in September 2020 was spatiotemporally correlated with the daily peak of hospitalizations during the second wave of the pandemic in November 2020 in France. During this period, an early first peak of daily reports of anosmia occurred among young adults on July 21, 2020, and the highest peak occurred on September 16, 2020, occurring 49 days before the peak of daily hospitalizations in November 2020. A reduction in the number of daily reports of anosmia by users of the web-based app also preceded the reduction in the number of daily hospitalizations by 10 and 9 days during the first and the second waves of the pandemic, respectively, in France.

During the first peak of hospitalizations in March 2020, we initially observed that the peak of daily reports of anosmia on our web-based app occurred 10 days before the peak in daily hospitalizations, which is similar to the mean 11-day period between infection and hospitalization among individuals experiencing severe forms of the disease as reported previously [6]. We did not assess the dynamics of RT–PCR tests in the population in March 2020 because RT–PCR tests were not performed on a large scale but were rather performed only for hospitalized individuals with COVID-19 at that time. The magnitude of connections to the web-based app was high in March 2020 because it occurred during the outset of media campaigns of the web-based app. No media campaigns occurred before the second wave of the pandemic in October 2020, which explains why the magnitude of the connections was lower than that during the first wave of the pandemic.

The second peak of daily reports of anosmia occurred in July 2020 without a subsequent increase in the number of hospitalizations. This may be explained by the young age of the infected users (median age 32 years), who were at a low risk of severe disease, and the summer season during which transmission to older individuals was limited by the elevated outdoor temperature and reduced indoor transmission. However, the reduction in the number of hospitalizations after the first wave of the pandemic decreased simultaneously, suggesting the initiation of a new outbreak in July 2020.

During the second wave of the pandemic in October 2020, the reduction in the positive outcomes on daily RT–PCR tests occurred only 2 days before the peaks in daily hospitalizations and ICU admission, suggesting a decreased potential of RT–PCR data to predict the precedence of the reduction in the number of hospitalizations when compared with the number of daily reports of anosmia. This issue was probably also associated with the delays in getting tested and in obtaining the results. Moreover, the analytic sensitivity of RT–PCR testing is high, and the long duration of the RNA-positive tail suggests that most infected individuals are being identified after the infectious period has passed. This may overestimate incidence of the disease and explain the lack of an association of daily testing with the peak in hospitalizations and ICU admission [7]. This overestimation is also associated with the high proportion of asymptomatic individuals identified through contact tracing investigations and self-testing of suspected individuals. The risk of saturation of biological laboratories and the lack of reactions on RT–PCR testing may also lead to the discordance of data from individuals testing positive from the actual dynamic of the outbreak, leading to the decreased potential of RT–PCR test data in predicting the hospitalization rate.

The potential of the web-based app as a predictor of the hospitalization rate is based on the high specificity of anosmia as a diagnostic feature of COVID-19, occurring few days before symptom exacerbation among hospitalized patients with COVID-19 [6].

We recorded a high peak of daily reports of anosmia on September 16, 2020, which was 49 days before the maximum peak of daily hospitalizations in November 2020, with a low median age of users (9.1% [n=190/2088] of whom were aged ≥45 years); this was followed by a second peak of hospitalizations among older individuals on October 26, 2020 (24.1% [n=402/1668] of whom were aged >45 years). The

anticipation of the peak in hospitalizations in these 49 days is probably specific to the summer period and subsequently to a relatively large first wave of infections among young adults—which occurred in mid-July 2020 without a significant increase in the hospitalization rate—and at a higher scale toward the end of summer vacations and during the back-to-school period after which many clusters of infection were observed at schools and universities. As young, internet-savvy users developed anosmia, they extensively used the app, but only few of them had severe COVID-19 and few required subsequent hospitalization. The September 2020 peak in daily reports of anosmia was probably followed by progressive disease transmission to older patients in October 2020, with a subsequent domino effect during low-temperature periods and increased indoor transmission. This may have triggered a second wave of infections among older users (who are at an increased risk of severe COVID-19) in October 2020, followed by a large wave of hospitalizations in November 2020. Since this app is used less extensively by older patients (in whom anosmia is less frequent), the magnitude of the daily reports of anosmia was lower in October 2020 than in September 2020, but the hospitalization rate increased. Thus, this app does not predict the magnitude of hospitalization after a peak in the reports of anosmia, unless focusing on older individuals.

During the first and the second wave of the pandemic in France, a reduction in the daily reports of anosmia preceded the reduction in the number of daily hospitalizations and ICU admissions by 10 and 9 days, respectively, in March and October 2020 but not in July or September 2020. This was not observed in July 2020 after a low hospitalization rate was reported among older individuals. However, daily hospitalizations stopped increasing in September 2020 a few days after the reduction in reports of anosmia in September 2020.

Although this tool does not accurately anticipate an increase in the magnitude of hospitalization, it seems to accurately predict the reduction in the hospitalization rate. The anticipation of the reduction in emergency department visits individuals reporting anosmia (by 7 days), hospitalizations (by 9 days), and ICU admissions (by 9 days) is thus crucial for anticipating the surge duration during the pandemic and in managing the requirements of beds in the ICU and dedicated COVID-19 wards.

We recorded more than 13,000,000 responses between March 17 and November 18, 2020, and 575,214 daily reports of anosmia from among an unknown number of users. As user data were anonymized, duplicate responses may have been obtained and were not assessable. However, the anonymous nature of the data collected from the app and the lack of tracking favored its extensive utilization in the French population, which was not observed with tracking applications.

As revealed from RT–PCR data, the infection rates decreased 15 days after large-scale curfew and lockdown measures were enforced, whereas daily reports of anosmia began decreasing 4 days later. This observation, along with spatiotemporal data obtained from reports of anosmia, can help the authorities nationwide to harness the value of emergency department visits, hospitalizations, and ICU admissions and to obtain early estimates of the outcomes of lockdown measures. Studies have assessed the generalizability of this tool in many countries in Europe (eg, UK and Germany) and in the United States [8-11].

Acknowledgments

We thank the app users for their participation in this study, as well as Magali Balavoine, MSc (Weprom, Angers, France). No participant was compensated for their contribution. We also thank the Caisse Primaire d'Assurance Maladie de Sarthe (Thibaut Lhermitte) and the French Ministry of Health (Dr Yann-Mael le Douarin), Docaposte (Denis Weiss, Denis De Amorim, Regis Senegou), and DernierCri. Weprom designed and conducted the study; collected, managed, analyzed, and interpreted the data; prepared, reviewed, and approved the manuscript; and approved the submission of this manuscript for publication. Weprom, FD, Docaposte, Kelindi, DernierCri, and provided administrative, technical, and material support.

Authors' Contributions

FD had complete access to all the study data and was responsible for the integrity of the data and the accuracy of the data analysis. FD conceptualized and designed the study. All authors acquired, analyzed, and interpreted the data, drafted and critically revised the manuscript for important intellectual content. FD performed the statistical analyses. FD, AF, and FXL supervised the study.

Conflicts of Interest

FD reports receiving a personal fee from AstraZeneca, Ipsen, Sivan Innovation, Kelindi, Pfizer, Chugai, and Roche. FD and FLG are the cofounders of Kelindi. SJ is the founder of Adobis Group. Outside of this work, AF receives fees from Gilead and MSD for educational seminars. FXL does not report any conflict of interest.

References

1. Denis F, Basch E, Septans A, Bennouna J, Urban T, Dueck AC, et al. Two-Year Survival Comparing Web-Based Symptom Monitoring vs Routine Surveillance Following Treatment for Lung Cancer. JAMA 2019 Jan 22;321(3):306-307 [FREE Full text] [doi: 10.1001/jama.2018.18085] [Medline: 30667494]
2. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. JAMA 2017 Dec 11;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]

3. Basch E. Patient-Reported Outcomes - Harnessing Patients' Voices to Improve Clinical Care. N Engl J Med 2017 Jan 12;376(2):105-108. [doi: 10.1056/NEJMp1611252] [Medline: 28076708]
4. Galmiche S, Rahbe E, Fontanet A, Dinh A, Bénézit F, Lescure F, et al. Implementation of a Self-Triage Web Application for Suspected COVID-19 and Its Impact on Emergency Call Centers: Observational Study. J Med Internet Res 2020 Nov 23;22(11):e22924 [FREE Full text] [doi: 10.2196/22924] [Medline: 33147165]
5. Denis F, Galmiche S, Dinh A, Fontanet A, Scherpereel A, Benezit F, et al. Epidemiological Observations on the Association Between Anosmia and COVID-19 Infection: Analysis of Data From a Self-Assessment Web Application. J Med Internet Res 2020 Jun 11;22(6):e19855 [FREE Full text] [doi: 10.2196/19855] [Medline: 32496206]
6. Spinato G, Fabbris C, Polesel J, Cazzador D, Borsetto D, Hopkins C, et al. Alterations in Smell or Taste in Mildly Symptomatic Outpatients With SARS-CoV-2 Infection. JAMA 2020 May 26;323(20):2089-2090 [FREE Full text] [doi: 10.1001/jama.2020.6771] [Medline: 32320008]
7. Mina MJ, Parker R, Larremore DB. Rethinking Covid-19 Test Sensitivity - A Strategy for Containment. N Engl J Med 2020 Nov 26;383(22):e120. [doi: 10.1056/NEJMp2025631] [Medline: 32997903]
8. Echeverría P, Mas Bergas MA, Puig J, Isnard M, Massot M, Vedia C, et al. COVIDApp as an Innovative Strategy for the Management and Follow-Up of COVID-19 Cases in Long-Term Care Facilities in Catalonia: Implementation Study. JMIR Public Health Surveill 2020 Jul 17;6(3):e21163 [FREE Full text] [doi: 10.2196/21163] [Medline: 32629425]
9. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul;26(7):1037-1040 [FREE Full text] [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
10. Urbach D, Awiszus F, Leiß S, Venton T, Specht AVD, Apfelbacher C. Associations of Medications With Lower Odds of Typical COVID-19 Symptoms: Cross-Sectional Symptom Surveillance Study. JMIR Public Health Surveill 2020 Dec 14;6(4):e22521 [FREE Full text] [doi: 10.2196/22521] [Medline: 33197879]
11. Zens M, Brammertz A, Herpich J, Südkamp N, Hinterseer M. App-Based Tracking of Self-Reported COVID-19 Symptoms: Analysis of Questionnaire Data. J Med Internet Res 2020 Sep 09;22(9):e21956 [FREE Full text] [doi: 10.2196/21956] [Medline: 32791493]

Abbreviations

ICU: intensive care unit

Edited by G Eysenbach; submitted 01.12.20; peer-reviewed by E Stavaux, D Urbach; comments to author 20.12.20; revised version received 04.01.21; accepted 01.03.21; published 12.03.21.

Please cite as:
Denis F, Fontanet A, Le Douarin YM, Le Goff F, Jeanneau S, Lescure FX
A Self-Assessment Web-Based App to Assess Trends of the COVID-19 Pandemic in France: Observational Study
J Med Internet Res 2021;23(3):e26182
URL: https://www.jmir.org/2021/3/e26182
doi:10.2196/26182
PMID:33709945

Original Paper

Effectiveness of Interactive Tools in Online Health Care Communities: Social Exchange Theory Perspective

Dixuan Ren[1], PhD; Baolong Ma[1], Prof Dr

School of Management and Economics, Beijing Institute of Technology, Beijing, China

Corresponding Author:
Baolong Ma, Prof Dr
School of Management and Economics
Beijing Institute of Technology
Number 5, Zhongguancun Road, Haidian District
Beijing
China
Phone: 86 1068915602
Email: dmmibb@163.com

Abstract

Background: Although the COVID-19 pandemic will have a negative effect on China's economy in the short term, it also represents a major opportunity for internet-based medical treatment in the medium and long term. Compared with normal times, internet-based medical platforms including the Haodf website were visited by 1.11 billion people, the number of new registered users of all platforms increased by 10, and the number of new users' daily consultations increased by 9 during the pandemic. The continuous participation of physicians is a major factor in the success of the platform, and economic return is an important reason for physicians to provide internet-based services. However, no study has provided the effectiveness of interactive tools in online health care communities to influence physicians' returns.

Objective: The effect of internet-based effort on the benefits and effectiveness of interactive effort tools in internet-based health care areas remains unclear. Thus, the goals of this study are to examine the effect of doctors' internet-based service quality on their economic returns during COVID-19 social restrictions, to examine the effect of mutual help groups on doctors' economic returns during COVID-19 social restrictions, and to explore the moderating effect of disease privacy on doctors' efforts and economic returns during COVID-19 social restrictions.

Methods: On the basis of the social exchange theory, this study establishes an internet-based effort exchange model for doctors. We used a crawler to download information automatically from Haodf website. From March 5 to 7, 2020, which occurred during the COVID-19 pandemic in China, cross-sectional information of 2530 doctors were collected.

Results: Hierarchical linear regression showed that disease privacy (β=.481; *P*<.001), reputation (β=.584; *P*<.001), and service quality (β=.560; *P*<.001) had a significant positive effect on the economic returns of the physicians. The influence of mutual help groups on earnings increases with an increase in the degree of disease privacy (β=.189; *P*<.001), indicating that mutual help groups have a stronger effect on earnings when patients ask questions about diseases regarding which they desire privacy.

Conclusions: For platform operators, the results of this study can help the platform understand how to improve doctors' economic returns, especially regarding helping a specific doctor group improve its income to retain good doctors. For physicians on the platform, this study will help doctors spend their limited energy and time on tools that can improve internet-based consultation incomes. Patients who receive internet-based health care services extract information about a doctor based on the doctor's internet-based efforts to understand the doctor's level of professionalism and personality to choose the doctor they like the most. The data used in this study may be biased or not representative of all medical platforms, as they were collected from a single website.

(J Med Internet Res 2021;23(3):e21892) doi:10.2196/21892

KEYWORDS

efforts; income; privacy disease; platform; social exchange

Introduction

Background

Since December 2019, the highly contagious novel coronavirus pneumonia (COVID-19) has been spreading within and beyond China [1]. The first case in China was identified in Wuhan City, Hubei Province, in early December [2]. To contain the outbreak, Wuhan imposed a quarantine from 10 AM on January 23, 2020, onward [3]. Most cities in Hubei Province were also quarantined, restricting people from traveling out of their own cities. People in proximity can easily transmit COVID-19; thus, social (ie, physical) distancing is an important measure to reduce its spread [4]. By February 2020, China had introduced travel bans and quarantine measures, closed public services, and canceled social events to contain COVID-19. COVID-19 also endangers people undergoing in-person care at hospitals as patients with different diseases gather in hospitals; therefore, the government stopped offline care to avoid spreading COVID-19 between doctors and patients [5]. However, medical consultations with patients who do not have COVID-19 remained a challenge during the pandemic.

The obvious answer to this challenge is online health care community (OHC), in which doctors can provide primary care for patients via internet-based platforms. This method can effectively prevent COVID-19 infections, which are deadly and highly transmissible [6,7]. Moreover, patients who do not have emergent diseases can receive doctor consultations. China Medical and Health Service System Planning (2015-2020) promulgated by the State Council notes that it is necessary to actively use the internet, cloud computing, and other information technologies to transform the existing health service model to benefit Chinese health service needs. Undoubtedly, the OHC will usher in unprecedented development opportunities as the government actively guides and supports the management of COVID-19 pandemic [8]. Compared with traditional offline health care, the OHC can prevent COVID-19 infections and operate outside of typical health care restrictions regarding time and space, as patients can communicate with doctors and obtain health information anytime and anywhere [6]. Therefore, the use of internet-based platforms to offer primary patient care provides a new and beneficial method to effectively allocate medical resources, improve doctor-patient relationships, and reduce medical costs and patient waiting time.

The prosperity of the OHC cannot be achieved without the continued participation and efforts of doctors, particularly high-status doctors who have always been a scarce medical resource. In the short term, doctors may want to gain reputation; however, in the long term, they still want to earn money [9]. Performance output is closely related to a person's effort at work [10]. The OHC provides convenience for patients and should consider the economic returns of doctors to encourage doctors to actively participate, change their methods, and reasonably allocate their own energy and time both online and offline. An effort has been defined as the accumulated amount of time invested, energy spent, or activity by which work is accomplished [10-12] and is composed of three factors: direction (working smart), level (working hard), and persistence (in terms of the amount of time spent working and continuing to try to achieve a goal in the face of failure). Doctors' efforts aim for patients to perceive the professional ability and attitude of service doctors, then patients feel satisfaction and hope for gaining further internet-based services, and the doctors' webpages are visited frequently, ultimately increasing the economic return of the doctors. The OHC mainly relies on the exchange of health knowledge and information between doctors and patients [13]. Therefore, the relationship between doctors' internet-based efforts and their economic returns is of great significance for OHC development. Whether health care workers can modify their traditional health care methods and actively participate in and adapt to the new health care model of the OHC remains to be determined. This study primarily reviews existing studies from the perspective of social exchange theory and the relationship between effort and earnings.

Literature Review

Social Exchanges Between Patients and Doctors

Social exchange theory describes the relationship among people from the perspective of benefits and costs. Individual behavior typically tries to maximize benefits by paying as little as possible [14], finding the optimal solution between costs and benefits. The participation of doctors in the OHC can be considered as a social exchange behavior that is described by social exchange theory [15] rather than a purely economic exchange. Social exchange theory advocates the use of economic methods to analyze noneconomic social behaviors and has been widely used to understand the dynamic exchange processes in social relationships [16]. On the basis of a systematic review of the literature on social exchange theory, 2 critical features of social exchange have been clarified: (1) dynamic interaction behaviors (ie, repeated exchange actions) and (2) the power in structural relationships (ie, the power to maintain the interdependent relationship among the exchangers in a social exchange) [17].

Relationship Between Efforts and Earnings

For service areas, efforts can enhance consumers' perceptions of the quality of products. Employees' work efforts influence customer satisfaction, and customers perceive that the harder employees work, the more satisfied they will be of the services they receive [18]. Efforts in the health consultation market are perceived as an exchange of resources in which doctors actively demonstrate their ability to attract more attention and obtain more consultation [18,19]: the harder employees work, the better their performance will be. There is a positive correlation between employee effort and job performance, and work effort level is one of the determinants of job performance [20,21]. Individuals or organizations with a strong sense of reciprocity are more willing to share knowledge and believe that knowledge sharing will yield certain personal rewards [22,23]. In the OHC market, there is a strong link between doctors' internet-based efforts and their internet-based rewards.

Research Model

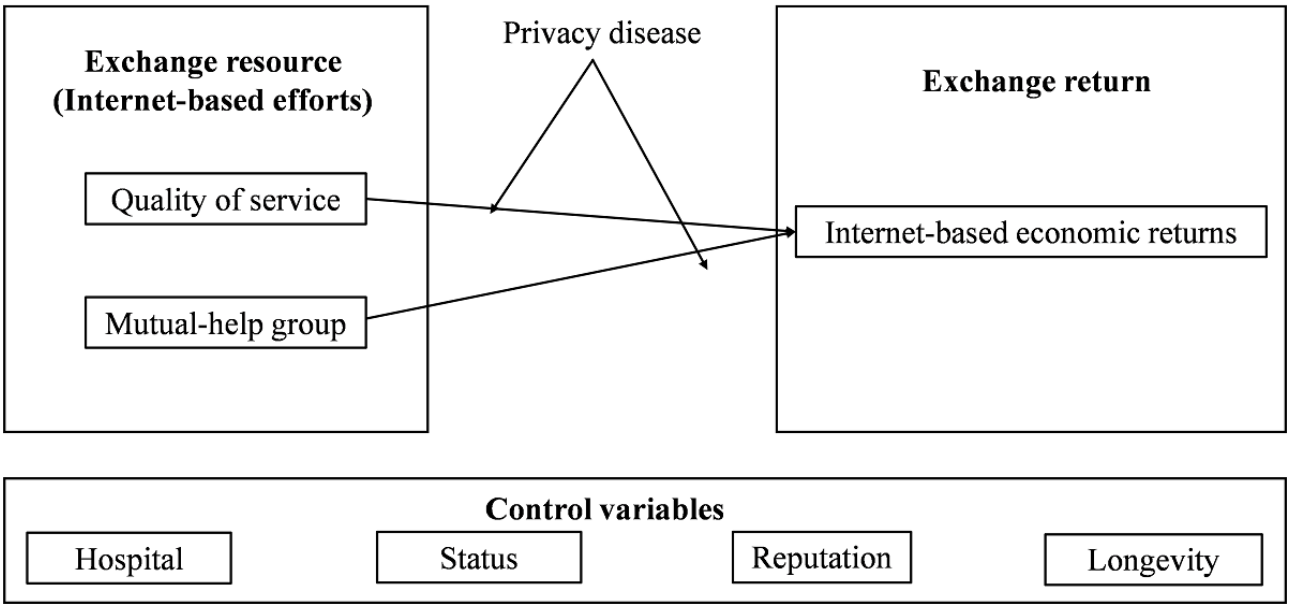
The professional capital used by doctors to participate in knowledge exchange describes the attitudes of interaction between the doctors and patients during social exchanges as well as the doctors' professional commitment (eg, enthusiasm,

responsibility, and morality) [24]. The professional capital of doctors can be directly observed by patients through doctors' effort behavior, which is shown in a series of dynamic interactive therapeutic behaviors because the interaction behavior of doctors is guaranteed by their ability to act independently and a set of rules of commitment [25]. As previously mentioned, studies have shown that effort has a positive effect on performance. The OHC is a typical information asymmetry market, and doctors know their professional abilities better than patients [9]. As recipients, patients interpret doctors' efforts and adjust their purchasing behavior accordingly. Patients who receive a doctor's effort are typically more likely to buy health care products than patients who do not receive it. On the basis of social exchange theory, doctors should provide information about their medical ability to users through effort behaviors to improve their earnings: the more doctors work, the more benefits they can obtain. The internet-based efforts of doctors can influence patient selection based on the social exchange function and affect the internet-based economic return and offline referrals of doctors. Patients believe that doctors who make more efforts possess higher service quality and therefore have a higher probability of being selected; thus, doctors gain higher economic returns when they put forth more effort.

Person efforts have not been considered in previous studies on the OHC market. There is a strong link between effort and performance, in which the OHC platform only supplies 2 interactive tools of effort between physicians and patients. However, effectiveness of interactive effort tools in internet-based health care areas remain unclear. Thus, we aim to (1) examine the effect of doctors' internet-based service quality on their economic returns during COVID-19 social restrictions, (2) examine the effect of mutual help groups built on their economic returns during COVID-19 social restrictions, and (3) explore the moderating effect of disease privacy on individual efforts and economic returns during COVID-19 social restrictions.

On the basis of social exchange theory, doctors' participation in the OHC is a process in which professional capital participates in social exchange for economic returns. The research model is illustrated in Figure 1.

Figure 1. Exchange model describing doctors' efforts and returns.



Methods

Study Design

In the OHC, patients can assess the value of doctors through the distribution and frequency of interactions with doctors (eg, the doctors' replies and the mutual help groups) to describe doctors' ability and willingness to work [26]. In this study, the benefit of internet-based consultation refers to the financial return that doctors obtain after performing certain tasks that expend time and energy. In combination with social exchange theory and the scenario investigated in this study, this study identifies two effort tools of doctor-patient interaction: quality of service and mutual help groups.

Interaction between sellers and consumers increases sellers' trust and promotes consumers' behavioral intentions [27]. Positive interactions that allow patients to feel the doctor's efforts and service attitudes online can help overcome distrust between doctors and patients.

OHCs are a collection of virtual discussion groups in which members can share feelings, knowledge, and experience about the health topics of their interests [28]. Doctors provide communication opportunities for patients with similar diseases. Patients can find valuable resources through a doctor's effort in a series of interactions in OHCs. Exchange behaviors are used to bridge doctors' efforts and patients' rewards.

Different disease types have different psychological characteristics, cognitive needs, and patient involvement [29]. Privacy diseases refer to diseases that patients are reluctant to disclose to others, as disclosure of relevant information about these diseases will lead to a series of consequences, such as patients' mental stress, love discrimination, disease discrimination, work pressure, and social stigma [16]. Thus,

OHCs provide a channel for privacy patients to acquire medical knowledge. Internet-based doctors' efforts (eg, the quality of service and mutual help groups) can effectively eliminate patients' concerns; thus, the positive effect on economic returns is stronger.

Thus, we hypothesize the following:

- Hypothesis 1 (H1): The service quality of physicians in OHCs is positively associated with the economic returns of doctors.
- Hypothesis 2 (H2): Mutual help groups in OHCs are positively associated with the internet-based economic returns of doctors.
- Hypothesis 3 (H3): The degree of disease privacy positively regulates the effect of physicians' service quality on revenue from internet-based consultations.
- Hypothesis 4 (H4): The degree of disease privacy positively regulates the influence of mutual help groups on the revenue of internet-based consultations.

Data Collection

This study collected data from haodf website for analysis. Haodf website is a typical internet-based health consulting service platform in China that has a long history of establishment, mature operation, high user coverage, high doctor activity, and high-quality hospital. Therefore, Haodf website was investigated in this study. By March 20, 2020, 591,832 doctors from 9614 hospitals had registered on the haodf website. Thus, haodf is a representative internet-based health advisory platform.

We used a crawler to download information automatically from Haodf website. From March 5, 2020, data collection lasted for 2 days and involved 2530 doctors. We collected the service quality, groups, status, and other relevant information for each doctor in the data set.

Measures

Physician engagement in OHCs is typically a form of social exchange with patients, which is characterized by structural power and dynamic interactions [15]. Income is important for social exchange. The product of the number of doctors' services and the unit price of consultation was used to determine the internet-based economic rewards.

Social exchange is a two-way transaction: "there must be both pay and return" [30]. Doctors can use two tools to show their efforts on the Haodf website: the doctors' quality of service and the mutual help group. The Haodf website provides patients with the ability to consult doctors [31]. This study uses the number of doctors who respond to user inquiries on the platforms as a variable of doctors' service quality. We assessed mutual help groups using the number of groups in which patients who consulted a doctor was assigned to different topic groups based on their disease characteristics. Communication in the same topic group occurred between the patients and doctors or patients with similar diseases. Doctors could provide service information and advice to the group. Disease privacy refers to a patient's reluctance to tell others or disclose that they have a disease [9]. Hepatitis B is highly contagious, and the disclosure of disease-related information may cause discrimination. Therefore, hepatitis B was investigated in this study as a disease with a high degree of privacy.

On the basis of the research model and hypotheses, the economic returns of internet-based care are affected by the internet-based efforts of doctors (quality of service and mutual help groups) and the degree of disease privacy. In addition, economic returns may be affected by other factors. First, the professional title of doctors affects doctors' internet-based economic returns. High-status doctors have higher priorities and privileges. For example, patients are always more willing to see the chief physician regardless of the doctors' professional level. Second, doctors' profits are affected by their reputation. Reputation in the internet-based market is primarily described through evaluation and feedback. Sellers with a better reputation have higher price premiums [32]. Finally, customers are more willing to visit century-old stores because they assume that stores with a longer operation time can deliver good service quality to customers. Similarly, if the doctor's home page has been operational longer and updated recently, the doctor's service quality is more likely to be recognized by patients. Therefore, city level, hospital level, doctor title, reputation, and internet-based service life was considered as control variables in this study.

Variables measured in this study are shown in Table 1.

Table 1. Measurement of variables.

Variables	Measurement
Dependent variables	
Internet-based economic rewards	The product of the number of doctors' services and the unit price of consultation.
Independent variables	
Doctors' quality of service	Sum up the number of doctors' answers from the patient consultation area of the doctor's home page.
Mutual help groups	Grab the count directly from the topic page of the doctor's personal home page.
Moderator variables	
Privacy disease	Hepatitis B
Control variables	
Hospital level	On the basis of the classification of 3 levels, 1 is the tertiary hospital, 2 is a secondary hospital, and 3 is the class-1 hospital. The higher the number, the higher the hospital's level.
Status	There are 5 grades: the chief physician is fifth grade, the associate chief physician is fourth grade, the attending physician is third grade, the resident physician is second grade, and other doctors are first grade. The higher the number, the higher the doctor's rank.
Reputation	As the value range of these 3 indicators, including votes, letter of thanks, and gifts, vary, they are averaged after standardization to measure the reputation of doctors.
Longevity	The number of years between the time the doctor's home page was created and the time the data were collected.

Statistical Results

Quantitative survey data were analyzed using SPSS (version 25.0; IBM Corporation). Table 2 shows the descriptive statistical results for all the variables. Associate chief physicians and chief physicians account for approximately half of all online doctors [33]. As shown in Table 2, the mean of doctor status was 4.52, indicating that the status of online doctors was above that of associated chief physicians. The average hospital rating was 2.99, indicating that the hospital where the online doctor worked offline was above level 2. Reputation is the average of the 3 indicators standardized (thank you letters, votes, and gifts); thus, its mean is 0. The average age of the doctor's home page was 8.46 years, the average number of online responses per doctor was 10,436.90, and the average number of mutual help groups built per doctor was 63.13.

Table 2. Results for descriptive statistics.

Variables	Value, mean (SD)	Minimum value	Maximum value
Returns	14,974.64 (32,179.683)	0	270,660
Quality	10,436.90 (10,427.042)	946	78,331
Groups	63.13 (152.230)	0	1651
Privacy	0.20 (0.399)	0	1
Hospital	2.99 (0.140)	1	3
Status	4.52 (0.721)	1	5
Reputation	0.15 (0.965)	−0.78	4.83
Reputation 1	120.20 (126.967)	0	75
Reputation 2	260.82 (254.238)	7	1421
Reputation 3	380.23 (466.302)	11	2785
Longevity	8.46 (2.419)	1.76	11.04

Results

To ensure that multicollinearity does not confound the regression results, this study first examined the variance inflation factors (VIFs) among variables. As all VIFs are lower than 5, no serious multicollinearity exists among the independent variables [34,35].

In this study, hierarchical regression analysis was used to examine the effect of the quality of service and mutual help groups on online economic rewards as well as the moderating effect of disease privacy. Considering the research design, the hierarchical regression analysis method is more suitable for this study because this study conducts a separate analysis of the variables of each layer to determine the differences among them. The purpose of the hierarchical regression analysis method is

to investigate the contribution of the variable to the regression equation when the contributions of other variables are excluded. The regression results of the hierarchical regression analysis in this study are presented in Table 3.

Table 3. Regression analysis.

Variables	Dependent variable: internet-based economic returns, nonstandardized coefficients (SE)		
	Model 1	Model 2	Model 3
Control variables			
Privacy	0.481[a] (0.127)	0.332[b] (0.114)	0.313[b] (0.108)
Hospital	−0.072 (0.362)	−0.008 (0.321)	−0.003 (0.299)
Status	0.087 (0.079)	0.124 (0.070)	0.120 (0.066)
Reputation	0.584[a] (0.054)	0.206[a] (0.067)	0.305[a] (0.065)
Longevity	−0.019 (0.024)	−0.024 (0.021)	−0.023 (0.020)
Independent variables			
Quality	N/A[c]	0.560[a] (0.069)	0.408[a] (0.069)
Groups	N/A	−0.062 (0.056)	−0.082 (0.052)
Quality × privacy	N/A	N/A	0.066 (0.047)
Groups × privacy	N/A	N/A	0.189[a] (0.053)

[a]*P*<.001.

[b]*P*<.01.

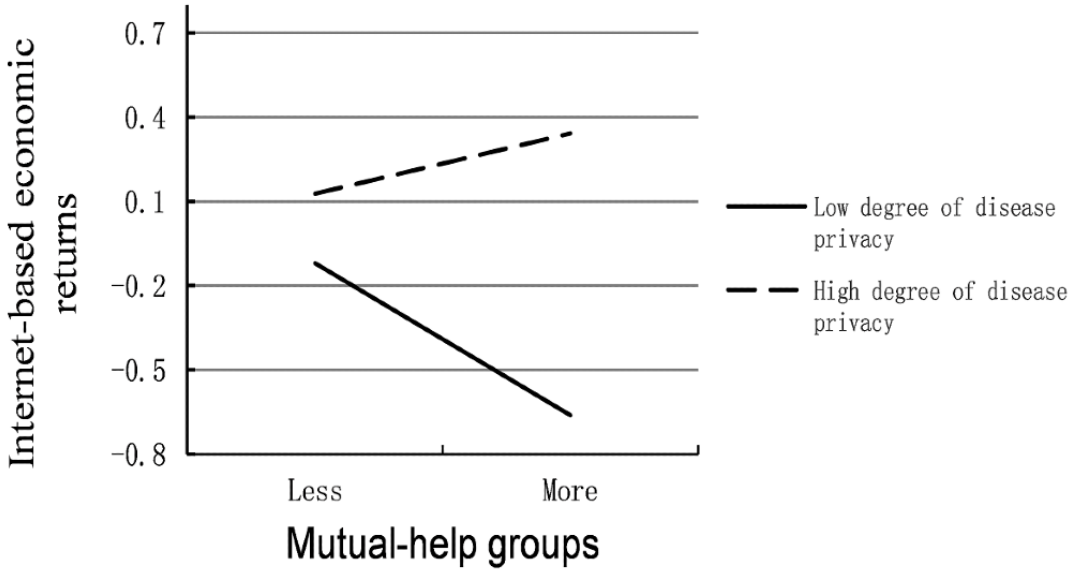
[c]N/A: not applicable.

Model 1 only contains control variables, and the results show that the degree of disease privacy (*P*<.001) and reputation (*P*<.001) have a significant positive effect; thus, doctors who provide medical services for patients with a high degree of disease privacy can obtain more economic returns. Channels between online and offline have substitutes and synergies [36,37]. Patients with nonprivacy diseases can go to hospitals without any concern, providing an alternative to internet-based medical services. However, patients with privacy diseases do not wish to go to hospitals for treatment and tend to use internet-based medical services. Therefore, OHCs tend to have patients with privacy diseases [9]. The internet-based reputation of doctors has a positive effect on doctors' internet-based incomes. A good reputation indicates that patients recognize the quality of service and indicates a lower risk of service quality. At a significance level of *P*<.05, the hospital level (*P*=.84), doctors' status (*P*=.27), and longevity (*P*=.42) were not significant predictors of internet-based economic returns. Model 1 explained 37% of the variation in dependent variables.

Model 2 added the influence of doctors' service quality and mutual help groups on the revenue of doctors' internet-based consultation based on model 1. According to the results shown in Table 3, *H1* doctors' service quality had a positive effect on internet-based economic returns, and thus *H1* was supported (*P*<.001). The *H2* mutual help group did not have a positive effect on internet-based consultation income (*P*=.27). The degree of interaction and the level of 2-way communication between the service provider and the patient are critical to the customer's ultimate perception of the service provider's service results [38]. Thus, doctors' replies are vital to patients as a result of these answers, which can alleviate distress. Model 2 explained 50.8% of the variation in the dependent variable. Compared with model 1, adding 2 tools of effort improved the explanatory ability of the model.

Finally, we analyzed the effects of the 2 interaction terms, which are the interaction terms between doctors' service quality and disease privacy, the interaction terms between mutual help group and disease privacy. The results indicate that *H3* is not supported (*P*=.16); however, *H4* is supported (*P*<.001). Privacy diseases can lead to social stigmatization, and these stigmatizations can have many negative effects on patients [39]. Although patients with privacy diseases are not willing to visit hospitals, they must receive doctors' care, the concern of other people, and health care information to recover. Model 3 explained 58% of the variation in the dependent variable, which indicates that compared with model 2, the 2 interaction terms increased the explanatory power of the model. The interaction between mutual help groups and disease privacy is shown in Figure 2. When the degree of disease privacy increases, the influence of mutual help group on earnings changes from a negative correlation to a positive correlation, indicating that when patients consult low disease privacy, the more mutual help group, the lower the incomes; however, when patients consult high disease privacy, the more mutual help group, the higher the incomes.

Figure 2. Interaction between mutual-help group and disease privacy.



Discussion

Principal Findings

The COVID-19 pandemic has significantly disrupted normal medical activities worldwide. To help decrease the spread of COVID-19, online health care platforms have been rapidly used by governments and users [40,41]. Currently, the Haodf website primarily provides 2 internet-based effort tools for doctors to interact with patients (doctors' quality of service and mutual help groups). In this study, we sought to examine the effect of these tools on the economic returns of doctors' internet-based consultations, and these mechanisms are adjusted by the degree of disease privacy. The primary findings of this study are summarized below.

Effect of Doctors' Internet-Based Service Quality on Their Economic Return

Doctors' internet-based service quality has a positive effect on the revenue of doctors' internet-based consultation, which indicates that the high doctors service quality to patients' consultation, the more economic return they obtain. In traditional Chinese culture, patients often judge the professional ability and service quality of doctors based on the hospital's grade and the doctor's status. Status is a sociological concept that focuses on rights or discrimination due to differences in social class rather than performance [42]. Patients have also become accustomed to long waiting times for outpatient registration, as Beijing has converged on the best resources in China. This situation causes many problems, including poor hospital environments and short communication between doctors and patients. Owing to COVID-19–related stay-at-home restrictions, in-person care is becoming less frequent. However, patients who do not leave home can use internet-based platforms for primary care in China. Thus, patients can save waiting time and enable better communication with high-status doctors from high-ranking hospitals. In addition, low-status doctors can put in their efforts in the OHC to earn more economic returns by patiently and carefully answering patients' inquiries.

Effect of Mutual Help Groups on Doctors' Economic Return

Mutual help groups did not affect their internet-based revenue; doctors could not obtain higher economic returns by building different types of topic groups. Whenever a given patient communicates with a doctor more than 3 times, the patient is automatically placed at the doctor's request into different virtual communities. Owing to COVID-19–related stay-at-home restrictions, in-person communication among patients is less frequent [43,44]. Doctors put patients with the same disease into a virtual group that provides convenient communication among its members. However, patients are not willing to participate in group discussions too much, as the members in the virtual community feel strange and lack trust in each other. Thus, mutual help groups do not affect the revenue of internet-based consultations.

The Moderating Effect of Disease Privacy

Disease privacy positively regulates the influence of mutual help groups on the revenue of internet-based consultation, which indicates that when the degree of disease privacy increases, mutual help groups increase the benefits of doctors' internet-based consultation. Patients with nonprivacy diseases easily communicate in person with others without any concerns. These patients can also communicate offline with patients who have similar diseases without the fear of discrimination or unfair treatment [45]; thus, they do not have strong intentions to find companions to communicate about their diseases in a virtual group. Patients with nonprivacy diseases in a virtual community were unfamiliar with each other, had a strong sense of strangeness, and could not trust each other. When patients have not joined a virtual community on their own [6], doctors' efforts would not be recognized by patients; therefore, the economic return would not increase. Conversely, patients with privacy diseases are reluctant to go to hospitals for consultation because of the risk to their privacy; thus, they have few opportunities to meet people who have a similar disease. In this situation, patients may not receive efficient or timely medical treatment because of a lack of medical knowledge and communication

with doctors. In virtual communities, doctors build a bridge between patients with privacy diseases and others in the outside world, and members in the mutual help group do not know each other; thus, users do not need to worry about disclosure of their disease status. Patients with privacy diseases were more willing to communicate with members of the group. Thus, doctors' efforts to build mutual help groups can be recognized by patients, which leads to improved economic returns.

Limitations

This study has produced some beneficial results; however, there are still certain limitations that must be addressed in future studies. First, data were taken from one OHC, the Good Doctor website, which is the most acceptable web-based consultation platform [33,46]; thus, the universality of the results is questionable. Future studies will consider collecting physician data from several different types of OHCs and empirically testing the model's findings. Second, in this study, only two tools (doctors' service quality and mutual help groups) provided by a single health consultation platform to describe the interaction efforts between doctors and patients were included, which may lead to an insignificant influence of the tools on the doctors' economic returns. In the future, the interactive efforts of other health consultation websites should be considered. Third, only one type of privacy disease was investigated (hepatitis B). Patients with different diseases have different degrees of involvement in information processing, and the degree of influence of information on patients' choices is also different [47-49]. Although the representativeness of this disease was considered in the selection of this study, further studies should consider diseases with different privacy levels, as different diseases that need privacy may have different service requirements.

Conclusions

In the face of the COVID-19 pandemic, internet-based medical treatment has become a new choice for people seeking medical treatment because of its unique features, including a lack of geographical restrictions, no in-person contact, and prevention of infection. Governments have rapidly used alternative methods for health care delivery, including web-based consultation, when people are restricted from their normal activities [5,8]. Many internet consultation platforms, including Haodf website, have seen a surge in visits during the epidemic. The prosperity of the platform depends on the continuous participation and efforts of doctors, and obtaining satisfactory economic returns is an important motivation for doctors' continuous participation. Therefore, it is of great significance for platforms, doctors, and patients to study internet-based consultation income.

This study expands and enriches the application and connotations of individual efforts. Individual effort has been focused on traditional industries and has never been used in online health care services. Online health services differ from traditional offline industries. This study expanded the application scope of efforts from the traditional offline industry to online health services and further identified 2 types of effort tools (doctors' replies and mutual help groups) based on the characteristics of the internet-based health market. These 2 effort tools are unique variables in the field of eHealth and have not been covered in previous studies in the field of electronic commerce. In addition, the degree of disease privacy is a unique variable in the field of OHCs, which has not been mentioned in previous studies in the field of electronic services, including electronic government and electronic commerce. This study identified that the uncertainty of privacy disclosure and the uncertainty of service quality may increase when the degree of disease privacy improves; thus, doctors must put in more efforts online to improve their services. Concurrently, this study also clearly distinguished the different influence mechanisms of the degree of disease privacy on the doctors' service quality and the mutual help groups. Doctors who built many mutual help groups could obtain higher incomes when serving patients with privacy diseases.

Acknowledgments

This research was supported by research funds of the National Natural Science Foundation of China (grant number: 71972012).

Conflicts of Interest

None declared.

References

1. Huang H, Peng Z, Wu H, Xie Q. A big data analysis on the five dimensions of emergency management information in the early stage of COVID-19 in China. J Chinese Gov 2020 Apr 21;5(2):213-233. [doi: 10.1080/23812346.2020.1744923]
2. Huang CL, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
3. Mu X, Yeh AG, Zhang X. The interplay of spatial spread of COVID-19 and human mobility in the urban system of China during the Chinese New Year. Environment and Planning B: Urban Analytics and City Science 2020 Sep 03:-. [doi: 10.1177/2399808320954211]
4. Ellis LA, Lee MD, Ijaz K, Smith J, Braithwaite J, Yin K. Covid-19 as 'Game Changer' for the physical activity and mental well-being of augmented reality game players during the pandemic: mixed methods survey study. J Med Internet Res 2020 Dec 22;22(12):e25117 [FREE Full text] [doi: 10.2196/25117] [Medline: 33284781]

5. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug 01;20(1):1193 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
6. Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents 2020 Mar;55(3):105924 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105924] [Medline: 32081636]
7. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med 2020 Apr 17;26(4):450-452 [FREE Full text] [doi: 10.1038/s41591-020-0820-9] [Medline: 32284615]
8. Isautier JM, Copp T, Ayre J, Cvejic E, Meyerowitz-Katz G, Batcup C, et al. People's experiences and satisfaction with telehealth during the covid-19 pandemic in Australia: cross-sectional survey study. J Med Internet Res 2020 Dec 10;22(12):e24531 [FREE Full text] [doi: 10.2196/24531] [Medline: 33156806]
9. Jia LJT, Ling J. Price premium research in the online health-care community. J Manag Sci 2018;31(1):15-32. [doi: 10.3969/j.issn.1672-0334.2018.01.002]
10. Rangarajan D, Jones E, Chin W. Impact of sales force automation on technology-related stress, effort, and technology usage among salespeople. Ind Market Manage 2005 May;34(4):345-354. [doi: 10.1016/j.indmarman.2004.09.015]
11. Mittal V, Ross WT, Tsiros M. The role of issue valence and issue capability in determining effort investment. J Mark Res 2018 Oct 10;39(4):455-468. [doi: 10.1509/jmkr.39.4.455.19122]
12. Srivastava R, Strutton D, Pelton LE. The will to win: an investigation of how sales managers can improve the quantitative aspects of their sales force's effort. J Mark Theory Pract 2015 Dec 08;9(2):11-26. [doi: 10.1080/10696679.2001.11501888]
13. Xiao N, Sharman R, Rao H, Upadhyaya S. Factors influencing online health information search: An empirical analysis of a national cancer-related survey. Decision Support Systems 2014 Jan;57:417-427. [doi: 10.1016/j.dss.2012.10.047]
14. Jin H. Empirical study of impacts of intrinsic and extrinsic motivations on employee knowledge sharing: crowding-out and crowding-in effect. J Manage Sci 2013;26(3):31-44. [doi: 10.3969/j.issn.1672-0334.2013.03.004]
15. Cropanzano R, Mitchell MS. Social exchange theory: an interdisciplinary review. J Manage 2016 Jul;31(6):874-900. [doi: 10.1177/0149206305279602]
16. Guo S, Guo X, Fang Y, Vogel D. How doctors gain social and economic returns in online health-care communities: a professional capital perspective. J Manage Info Sys 2017 Aug 17;34(2):487-519. [doi: 10.1080/07421222.2017.1334480]
17. Choi BCF, Jiang Z, Xiao B, Kim SS. Embarrassing exposures in online social networks: an integrated perspective of privacy invasion and relationship bonding. Info Sys Res 2015 Dec;26(4):675-694. [doi: 10.1287/isre.2015.0602]
18. Chai S, Bagchi-Sen S, Morrell C, Rao HR, Upadhyaya SJ. Internet and online information privacy: an exploratory study of preteens and early teens. IEEE Trans Profess Commun 2009 Jun;52(2):167-182. [doi: 10.1109/tpc.2009.2017985]
19. Ahearne M, Rapp A, Hughes DE, Jindal R. Managing sales force product perceptions and control systemsinthe success of new product introductions. J Mark Res 2010 Aug;47(4):764-776. [doi: 10.1509/jmkr.47.4.764]
20. Kacmar KM, Zivnuska S, White CD. Control and exchange: The impact of work environment on the work effort of low relationship quality employees. Leadersh Q 2007 Feb;18(1):69-84. [doi: 10.1016/j.leaqua.2006.11.002]
21. Jia BT, Zhong P. An analysis of the impact of positive feedback from customers on effort intention of frontline employees. J Southwest Jiaotong Uni 2010;11(2):77-81. [doi: 10.35741/issn.0258-2724]
22. Sun SJ, Chumg H. Influence on willingness of virtual community's knowledge sharing: based on social capital theory and habitual domain. Worl Aca Sci 2009;53:142-149. [doi: 10.5281/zenodo.1085155]
23. Hannan RL. The combined effect of wages and firm profit on employee effort. Account Rev 2005;80(1):167-188. [doi: 10.2308/accr.2005.80.1.167]
24. LaClair M. Professional capital: transforming teaching in every school. J Edu Pol 2015 Feb 11;30(3):464-466. [doi: 10.1080/02680939.2015.1008741]
25. Elizabeth B. Building professional capital: New Zealand social workers and continuing education. Melbourne: Deakin University; 2010. URL: http://dro.deakin.edu.au/view/DU:30032402 [accessed 2021-02-10]
26. Rincon-Gallardo S, Fullan M. Essential features of effective networks in education. J Prof Cap Community 2016;1(1):2016-2022. [doi: 10.1108/jpcc-09-2015-0007]
27. Yu XZ, Guo WJ. Research on the influence of online product information presentation on customers? Behaviour intention. Inf Stud: Theory Appl 2017:80-84. [doi: 10.16353/j.cnki.1000-7490.2017.10.015]
28. Yan L, Tan Y. Feeling blue? Go online: an empirical study of social support among patients. Info Sys Res 2014 Dec;25(4):690-709. [doi: 10.1287/isre.2014.0538]
29. Jin J, Yan X, Li Y, Li Y. Int J Med Inform 2016 Feb;86:91-103. [doi: 10.1016/j.ijmedinf.2015.11.002] [Medline: 26616406]
30. Molm LD. Theoretical comparisons of forms of exchange. Sociol Theory 2016 Jun 24;21(1):1-17. [doi: 10.1111/1467-9558.00171]
31. Wu H, Lu N. Service provision, pricing, and patient satisfaction in online health communities. Int J Med Inform 2018 Feb;110:77-89. [doi: 10.1016/j.ijmedinf.2017.11.009] [Medline: 29331257]
32. Ba SL, Pavlou PA. Evidence of the effect of trust building technology in electronic markets: price premiums and buyer behavior. MIS Quarterly 2002 Sep;26(3):243. [doi: 10.2307/4132332]

33. Li Y, Yan X, Song X. Provision of paid web-based medical consultation in China: cross-sectional analysis of data from a medical consultation website. J Med Internet Res 2019 Jun 03;21(6):e12126 [FREE Full text] [doi: 10.2196/12126] [Medline: 31162129]
34. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: indeed a silver bullet. J Mar Theory Pract 2014 Dec 08;19(2):139-152. [doi: 10.2753/MTP1069-6679190202]
35. Belsley DK, Welsch R. Regression diagnostics: identifying influential data and sources of collinearity. Hoboken, NJ: John Wiley & Sons; 2004.
36. Antoun J. Electronic mail communication between physicians and patients: a review of challenges and opportunities. Fam Pract 2016 Apr 28;33(2):121-126. [doi: 10.1093/fampra/cmv101] [Medline: 26711957]
37. Wu H, Lu N. Online written consultation, telephone consultation and offline appointment: an examination of the channel effect in online health communities. Int J Med Inform 2017 Nov;107:107-119. [doi: 10.1016/j.ijmedinf.2017.08.009] [Medline: 29029686]
38. Dagger TS, Sweeney JC, Johnson LW. A hierarchical model of health service quality. J Ser Res 2016 Jun 29;10(2):123-142. [doi: 10.1177/1094670507309594]
39. Major B, O'Brien LT. The social psychology of stigma. Annu Rev Psychol 2005 Feb;56(1):393-421. [doi: 10.1146/annurev.psych.56.091103.070137] [Medline: 15709941]
40. Thomas EE, Haydon HM, Mehrotra A, Caffery LJ, Snoswell CL, Banbury A, et al. Building on the momentum: sustaining telehealth beyond COVID-19. J Telemed Telecare 2020 Sep 26:-. [doi: 10.1177/1357633x20960638]
41. Snoswell CL, Caffery LJ, Haydon HM, Thomas EE, Smith AC. Telehealth uptake in general practice as a result of the coronavirus (COVID-19) pandemic. Aust Health Rev 2020;44(5):737. [doi: 10.1071/ah20183]
42. Washington M, Zajac EJ. Status evolution and competition: theory and evidence. Aca Manage J 2005 Apr;48(2):282-296. [doi: 10.5465/amj.2005.16928408]
43. Wilhelm JA, Helleringer S. Utilization of non-Ebola health care services during Ebola outbreaks: a systematic review and meta-analysis. J Glob Health 2019 Jun;9(1):- [FREE Full text] [doi: 10.7189/jogh.09.010406] [Medline: 30701070]
44. Wang SY, Chen LK, Hsu SH, Wang SC. Health care utilization and health outcomes: a population study of Taiwan. Health Policy Plan 2012 Oct;27(7):590-599. [doi: 10.1093/heapol/czr080] [Medline: 22258470]
45. Hong W, Chan FKY, Thong JYL, Chasalow LC, Dhillon G. A framework and guidelines for context-specific theorizing in information systems research. Info Sys Res 2014 Mar;25(1):111-136. [doi: 10.1287/isre.2013.0501]
46. IResearch. Report of China internet medical doctors' needs in 2016. URL: http://report.iresearch.cn/report/201609/2650 [accessed 2021-02-10]
47. Hu X, Wu G, Wu Y, Zhang H. The effects of Web assurance seals on consumers' initial trust in an online vendor: a functional perspective. Deci Sup Sys 2010 Jan;48(2):407-418. [doi: 10.1016/j.dss.2009.10.004]
48. Utz S, Kerkhof P, van den Bos J. Consumers rule: How consumer reviews influence perceived trustworthiness of online stores. Elec Comm Res and Appl 2012 Jan;11(1):49-58. [doi: 10.1016/j.elerap.2011.07.010]
49. Wen J, Lin Z, Liu X, Xiao SH, Li Y. The interaction effects of online reviews, brand, and price on consumer hotel booking decision making. J Trav Res 2020 Apr 04:-. [doi: 10.1177/0047287520912330]

Abbreviations

OHC: online health care community
VIF: variance inflation factor

Edited by G Eysenbach; submitted 28.06.20; peer-reviewed by C Miranda, D Bowen; comments to author 17.11.20; revised version received 12.01.21; accepted 22.01.21; published 12.03.21.

Please cite as:
Ren D, Ma B
Effectiveness of Interactive Tools in Online Health Care Communities: Social Exchange Theory Perspective
J Med Internet Res 2021;23(3):e21892
URL: https://www.jmir.org/2021/3/e21892
doi:10.2196/21892
PMID:33709940

Viewpoint

Artificial Intelligence–Aided Precision Medicine for COVID-19: Strategic Areas of Research and Development

Enrico Santus[1,2], PhD; Nicola Marino[2,3], BSc; Davide Cirillo[2,4], PhD; Emmanuele Chersoni[5], PhD; Arnau Montagud[4], PhD; Antonella Santuccione Chadha[2,6], MD; Alfonso Valencia[4,7], PhD; Kevin Hughes[8], MD; Charlotta Lindvall[9,10], MD, PhD

[1]Division of Decision Science and Advanced Analytics, Bayer Pharmaceuticals, Whippany, NJ, United States
[2]The Women's Brain Project, Zurich, Switzerland
[3]Department of Medical and Surgical Sciences, Università degli Studi di Foggia, Foggia, Italy
[4]Barcelona Supercomputing Center, Barcelona, Spain
[5]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)
[6]Biogen International GmbH, Baar, Switzerland
[7]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain
[8]Massachusetts General Hospital, Boston, MA, United States
[9]Dana-Farber Cancer Institute, Boston, MA, United States
[10]Harvard Medical School, Boston, MA, United States

Corresponding Author:
Davide Cirillo, PhD
Barcelona Supercomputing Center
c/Jordi Girona, 29
Barcelona
Spain
Phone: 34 934137971
Email: davide.cirillo@bsc.es

Abstract

Artificial intelligence (AI) technologies can play a key role in preventing, detecting, and monitoring epidemics. In this paper, we provide an overview of the recently published literature on the COVID-19 pandemic in four strategic areas: (1) triage, diagnosis, and risk prediction; (2) drug repurposing and development; (3) pharmacogenomics and vaccines; and (4) mining of the medical literature. We highlight how AI-powered health care can enable public health systems to efficiently handle future outbreaks and improve patient outcomes.

(J Med Internet Res 2021;23(3):e22453) doi:10.2196/22453

KEYWORDS

COVID-19; SARS-CoV-2; artificial intelligence; personalized medicine; precision medicine; prevention; monitoring; epidemic; literature; public health; pandemic

Introduction

The ongoing COVID-19 pandemic has highlighted the fragility of the health care system during unexpected events, testing the endurance of even the top-performing ones [1]. As noted by several scholars, embracing artificial intelligence (AI) for health care optimization and outcome improvement is not an option anymore [2]. Concerning the ongoing COVID-19 pandemic, several studies have highlighted that the timely inclusion of AI-powered technologies would have accelerated the identification of and effective response to COVID-19 outbreaks worldwide. An example is the widely reported algorithm from the Canadian company BlueDot, based on natural language processing (NLP) and machine learning, which forecasted the emerging risk of a virus spread in Hubei province in late December 2019, by screening news reports and airline ticketing [3].

Awareness of the benefits of employing AI to support and manage the COVID-19 crisis and its aftermath is increasing, particularly in the medical and research community. Notable examples of early AI-powered contributions include the discovery of relevant SARS-CoV-2 target proteins by

DeepMind's AlphaFold [4] and the design by Infervision of a computer vision algorithm for the detection of coronavirus pneumonia based on lung images [5].

Benefits do, however, come with technical challenges and related risks that still need to be properly assessed. For example, the absence of transparency and interpretability in AI models obscures the fact that the efficacy of these technologies is not equal across population groups. COVID-19 incidence and outcomes vary according to a large number of individual factors, including age, sex, ethnicity, health status, drug utilization, and others [6]. Sensitizing AI technologies to the diversity of the patient population and ensuring data security [7] is imperative to avoid biased decisions [8-10]. Therefore, a crucial step to obtain robust, trustworthy, and intelligible applications that account for demographic equity is to assess potential biases in the resources used to train AI models for precision medicine [11].

As of today, AI systems are, regrettably, rarely endowed with robustness to class imbalances, such as sex and gender groups [12]. In this regard, sex differences in COVID-19 cases, as well as sex-specific risk factors and socioeconomic burden, have been recently highlighted in a case study by the European Commission [13]. Dataset multidimensionality that can fairly represent the population constitutes one of the main challenges for biobanking and cohort design efforts that collect different axes of health data [14]. In this regard, fair and broad data collection systems are of primary importance. Two essential international references for COVID-19 genomic and medical data are the EMBL-EBI COVID-19 Data Portal [15] and the NIH National COVID Cohort Collaborative (N3C) [16]. The COVID-19 Host Genetics Initiative [17] is an international collaborative undertaking to share resources to investigate the genetic determinants of COVID-19 susceptibility, severity, and outcomes [18]. The Coronavirus Pandemic Epidemiology (COPE) consortium aims to involve experts in the development of a personalized COVID-19 Symptom Tracker mobile app that works as a real-time data capture platform [6], which garnered over 2.8 million users in a few days. Moreover, COVID-19 sex-disaggregated data are collected by Global Health 50/50 [19], an initiative housed at University College London, advocating for gender equity.

Other ethical concerns include life-or-death decisions through risk prediction models, which may help optimize resource allocation in times of scarcity. The application of nonoptimal models may incur the risk of worsening biases and exacerbating disparities for people with serious illnesses and different treatment priorities, potentially causing the reduction in the use of services rather than achieving the best patient care [20]. Nevertheless, the power of prediction models is impressive, and it may play a key role in the future if properly exploited. For instance, a study from Cambridge University [21] shows how the use of secure AI operating on anonymized COVID-19 data can accurately predict the patient journey, allowing an optimal allocation of resources and enabling well-informed and personalized health care decision-making. This is a particularly important point, especially considering the difficulty in managing the increasing need for intensive care units (ICUs) during the COVID-19 pandemic peak [22,23].

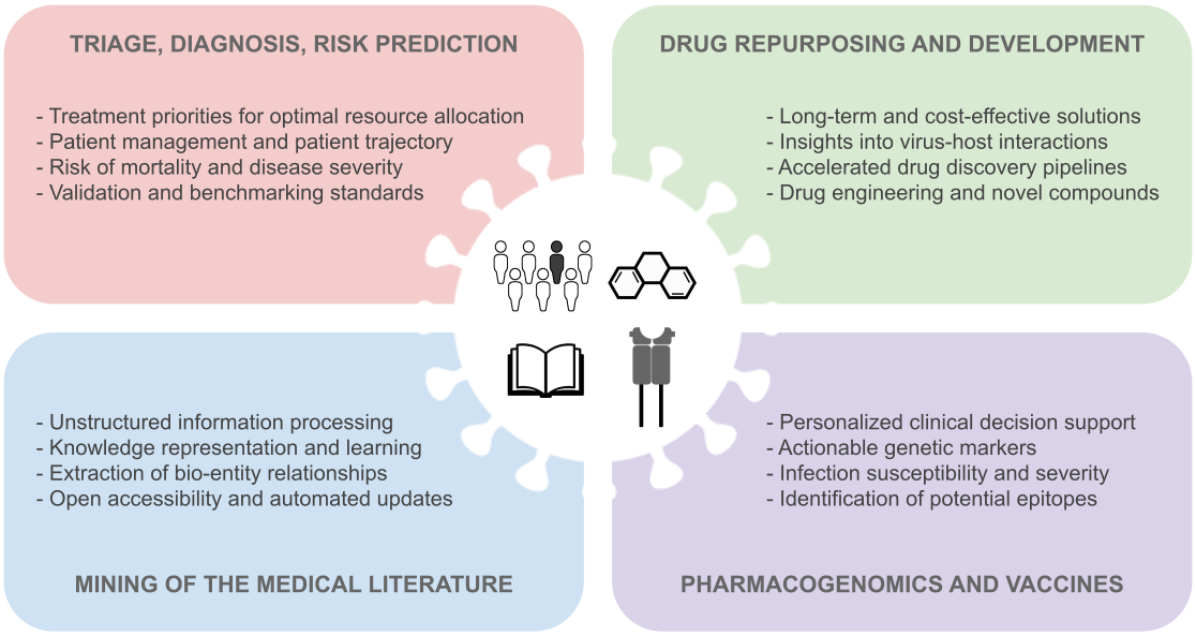
The way the AI systems will be exploited is probably the most delicate topic in this adoption process, particularly if we refer to the decisional independence of the medical staff. As humans, in fact, clinicians are also affected by numerous cognitive biases, including the *confirmation bias*, which may lead them to give excessive importance to the evidence supporting automated prediction (eg, risk prediction, diagnosis, and treatment suggestion) and ignore evidence that refutes it [8,24].

Despite the abovementioned concerns, there are numerous success stories in the adoption of risk prediction models. For example, Duke University adopted a system called Sepsis Watch that identifies in advance the inflammation leading to sepsis—one of the leading causes of hospital deaths. Within two years from the tool introduction, the number of sepsis-induced patients drastically decreased [25], thanks to three key elements: (1) adaptation of the predictive model to a highly specific context; (2) scalability through integration with hospital workflows; and (3) the adopted user experience–based approach, which places clinicians and health care professionals at the center of the software development process, adhering with the human-in-the-loop paradigm [26,27].

The COVID-19 crisis is accelerating anticipated changes towards a stronger collaboration between computer science and medicine. In particular, the crisis has exposed the need for increased scrutiny of the relationship between AI and patients as well as health care personnel under the lens of human and emotional needs, as demonstrated by the surge of mental health consequences of the pandemic [28] and the growing development of AI-based mental health apps and related digital tools [29]. Such aspects, together with others related to general data access and the use of AI for disease outcome prediction, are fueling the current debate about the convergence of AI and medicine [30,31] and the actionable realization of AI-powered innovations to bridge the gap between technological research and medical practice, including applications in medical triage and advice, diagnostics and risk-adjusted paneling, population health management, and digital devices integration [32]. Concerning this aspect, it is important to mention the recent publication of guidelines for the rigorous and transparent adoption of AI in the clinical practice: CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) [33] and SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence) [34].

Translating patient data to successful therapies is the major objective of implementing AI for health [35], especially in times of a pandemic crisis, with the ultimate goal of achieving a successful bench-to-bedside model for better clinical decision-making [36,37]. In this work, we review some major examples of what AI has achieved during the COVID-19 pandemic and the challenges that this technology and the medical community are currently facing in four main strategic areas of research and development (Figure 1): (1) triage, diagnosis, and risk prediction; (2) drug repurposing and development; (3) pharmacogenomics and vaccines; (4) mining of the medical literature.

Figure 1. Main strategic areas of research and development for the realization of artificial intelligence (AI) to fight COVID-19: (1) triage, diagnosis, and risk prediction; (2) drug repurposing and development; (3) pharmacogenomics and vaccines; and (4) mining of the medical literature. The text within the four panels enlists the advantages and actionable solutions exhibited by the AI-aided precision medicine approaches surveyed in this work.



Triage, Diagnosis, and Risk Prediction

AI has been applied to determine treatment priorities in patients with COVID-19 or triage and to better allocate limited resources. A group of researchers at the General Hospital of the People's Liberation Army (PLAGH), Beijing, China, has developed an online triage tool model [38] to manage suspected COVID-19 pneumonia in adult patients with fever [39]. Using clinical symptoms, routine laboratory tests, and other clinical information available at admission (eg, clinical features), they trained a model based on logistic regression with the least absolute shrinkage and selection operator (LASSO), obtaining an area under the receiver operating characteristic curve (AUROC) of 0.841 (100% sensitivity and 72.7% specificity). Based on data from two hospitals in Wenzhou, Zhejiang, China, another study group recently used an entropy-based feature selection approach: they modeled combinations of clinical features that could identify initial presentation patients who are at a higher risk of developing severe illness, with an accuracy of 80% [40]. Their results show that mildly elevated alanine aminotransferase levels, the presence of myalgias (body aches), and an elevated hemoglobin level (red blood cells), in this order, are predictive of the later development of acute respiratory distress syndrome.

A thorough study on risk prediction was carried out at the University of Cambridge based on the development of a proof of concept system to model the full patient journey through risk prediction models [21]. By identifying the risk of mortality and ICU/ventilator need, the system aims at enabling doctors to answer questions such as: Which patients are most likely to need ventilators within a week? How many free ICU beds in the hospital are we likely to have in a week from now? Which of two patients will get more benefits from going on a ventilator today? The predictive models showed accuracies ranging from 77% for ventilator need to 83% for ICU admission and 87% for mortality.

Risk prediction models are not new to the AI-aided health care approach. They have already been successfully utilized for tasks such as predicting the risk of developing cancer [41,42] and identifying which patients are likely to benefit from heart-related procedures [43]. However, the COVID-19 crisis has accelerated the utilization of such models. In a recent study, Wynants and collaborators [44] screened 14,217 published titles about the pandemic from PubMed and Embase (Ovid, arXiv, medRxiv, and bioRxiv), finding over 107 studies describing 145 prediction models. Among them, 4 models aimed to identify people at risk and 50, to predict the mortality risk, progression to severe disease, ICU admission, ventilation, intubation, or length of hospital stay. These models not only provide interesting results but also inform about the most valuable predictors, such as age, body temperature, lymphocyte count, and lung imaging features. Despite this, these models cannot be directly applied in the clinical setting without further validation, in order to guarantee data and experiment transparency and robustness, together with decision interpretability and model generalizability.

The remaining 91 models from this study were dedicated to the diagnosis of COVID-19, 60 of which exploited medical imaging. This number clearly shows that diagnosis is another important field for the application of AI techniques [45], with digital pathology exhibiting high effectiveness. In particular, convolutional neural networks (CNNs) have been supporting radiologists in their expert decisions [46]. In a recent study, a CNN was trained to automatically learn patterns related to COVID-19 (ie, ground-glass opacities, multifocal patchy

consolidation, and/or interstitial changes with a predominantly peripheral distribution), achieving an AUROC of 0.996 (98.2% sensitivity and 92.2% specificity) and outperforming the reverse-transcription polymerase chain reaction, which also suffers from a significant time lag. In addition to accuracy, these approaches put the speed of the diagnosis on the table: CNNs can analyze up to 500 images in a few seconds, whereas radiologists would need hours to complete the same task.

Although chest computed tomography (CT) scans represent a commonly exploited source of information to train AI to rule out SARS-CoV-2 infection, the rapid detection of patients with COVID-19 can greatly benefit from learning approaches that utilize heterogeneous types of data. In this regard, it is crucial to consider the importance of training CNNs in a correct gender balance in medical imaging datasets to avoid producing distorted classifications for assisted diagnosis [12]. Moreover, it is crucial to rely on high-quality benchmarking and robust validation strategies to assess the generalization of the model to other datasets and populations [47,48].

Indeed, AI can exploit multidimensional data, including the series of epidemiological, clinical, biological, and radiological criteria defined by the World Health Organization [49]. In a collaboration between researchers at hospitals in China and in the USA, CNN and other machine learning methods (eg, support vector machine, random forest, and neural networks) have been used to model and integrate CT scans and clinical information for diagnostic purposes [45]. The joint model that uses both information sources achieved a 0.92 AUROC (84.3% sensitivity and 82.8% specificity), outperforming the individual models. Moreover, the models allowed the identification of age, viral exposure, fever, cough, cough with sputum, and white blood cell counts as the main features associated with SARS-CoV-2 infection status.

Recently, the National Institute of Biomedical Imaging and Bioengineering has launched the Medical Imaging and Data Resource Center with the goal of coupling AI and medical imaging for COVID-19 early detection and personalized therapies [50].

AI has also been utilized to identify patients at higher risk of mortality. Researchers at the Tongji Hospital, Wuhan, China, have screened electronic health records of 375 discharged patients to use clinical measurements as features and have trained a gradient-boosted decision tree model to predict mortality risk [51]. The accuracy of the system was 93%. Its utilization would make it possible for physicians to immediately identify critical cases and act accordingly. The model was also able to detect three key clinical features, that is, lactic dehydrogenase, lymphocyte count, and high-sensitivity C-reactive protein.

Drug Repurposing and Development

Although triage, diagnosis, and risk prediction are three of the most relevant tasks that AI has helped with during the peaks of the pandemic, other objectives are currently being addressed for long-term solutions. Among them are target selection for drug repurposing [52] and approaches for drug development, including de novo drug design [53].

Drug repurposing comprises identifying existing drugs that could effectively act on proteins targeted by the virus. Recently, 332 high-confidence SARS-CoV-2 protein–human protein interactions have been experimentally identified, as well as 69 ligands, comprising drugs approved by the US Food and Drug Administration (FDA) and compounds in preclinical and clinical trials, which specifically target these interactions [54]. Understanding which proteins and pathways in the host the virus targets during infection is crucial for the development of AI systems for drug repurposing.

For instance, algorithms modeling the interaction between drugs and proteins have helped identify baricitinib, which was previously used for the treatment of arthritis, as a useful drug against COVID-19 [55]. This drug inhibits the proteins that help the virus penetrate the host cell. Thanks to approaches that exploit the computational identification of relations between existing drugs and target molecules, research published by a team of Korean and American scientists has allowed the identification of FDA-approved antivirals that could potentially target the key proteins for COVID-19 [56].

The molecular processes of virus-host interactions have been recently reconstructed in an international effort coordinated by domain experts, called the COVID-19 Disease Map project [57]. The project aims to maintain an open-access resource for continuous, curated integration of data and knowledge bases to support computational analysis and disease modeling. It represents a milestone of paramount importance for the development of AI systems for SARS-CoV-2 and their comparison with models of other coronaviruses. Moreover, by providing information about the intermolecular wiring of virus-host interactions, the project enables network-based AI modeling for COVID-19 drug repurposing, which has recently shown promising results by using network diffusion and network proximity [58]. Moreover, deep neural networks largely employed in NLP, such as the Transformer architecture, have also been proposed for COVID-19 drug repurposing [56].

In the field of drug development, that is, the pharmacotherapeutic course of a newly identified lead compound, computational models have been proven extremely successful in facilitating a quicker, cheaper, and more effective development of new drugs [59]. For instance, AI can map multidimensional characteristics of proteins to considerably speed up the research process in comparison to traditional methodologies such as x-ray crystallography. In this regard, AI is crucial in optimizing drug discovery pipelines and improving drug development outcomes, with estimated costs of US $2.6 billion [59].

Structural modeling and chemoinformatics methods for COVID-19 (eg, docking-based binding conformation studies of small molecules to target human or viral proteins) can greatly benefit from AI solutions. For instance, AI-based approaches have been used to infer structural similarities among molecules, such as algorithms that can model the graphical structure of chemical compounds through graph convolutional networks or other approaches [60]. AI systems can also leverage knowledge

about protein sequences to infer the activity of similar ones. As previously mentioned, Google DeepMind has managed to predict the structure of five proteins targeted by SARS-CoV-2, namely SARS-CoV-2 membrane protein, Nsp2, Nsp4, Nsp6, and papain-like proteinase (C-terminal domain) [4]. The deep learning approach uses amino acid features from similar sequences, based on multiple sequence alignment, to infer the distribution of structural distances to predict the protein structures [61].

Finally, AI can also be used to synthetically generate new molecules, such as new chemical compounds. For instance, the biotech company Insilico Medicine used reinforcement learning to model small molecules and identify those that inhibit specific infection pathways. The team created a generative chemistry pipeline to design novel SARS-CoV-2 inhibitors to later be synthesized and tested. The pipeline employs a large array of generative models, including autoencoders, generative adversarial networks, and genetic algorithms optimized with reinforcement learning [53].

Pharmacogenomics and Vaccines

Pharmacogenomics, which is the study of the role of genomic characteristics of an individual in drug response, represents a key gateway to personalized medicine [62-64]. Although the translation of genomic information into clinical practice is recognized as one of the most challenging aspects of the future of medicine [65], the information about the genetic makeup of individual patients has the potential to guide clinical decision support and to facilitate biomedical research in many different areas. For instance, genomics can inform drug discovery by providing simultaneous insights into the disease mechanisms and potential targets for treating individual patients [66].

Pharmacogenomics approaches to COVID-19 are still in their infancy. Indeed, although the SARS-CoV-2 genome was published in draft on January 10, 2020 [67], and real-time tracking of the pathogen evolution is now available [68], much less genomic information is currently available about the host. Several studies focus on genetic variations associated with susceptibility to infection and clinical manifestations, including human leukocyte antigen (HLA) variants in the UK Biobank population-based cohort [69] and angiotensin-converting enzyme 2 (ACE2) variants in the Italian population [70]. Retrospective and prospective studies focusing on COVID-19 disease susceptibility and severity have been collected by the COVID-19 Host Genetics Initiative [17,18].

Despite the absence of direct evidence of pharmacogenomics data in COVID-19 patients, the related literature for COVID-19 therapies, including hydroxychloroquine, ribavirin, and baricitinib, has been recently surveyed [71]. Potential actionable genetic markers have been reported, namely, several genetic variants that can alter the pharmacokinetics of drugs that may affect the response to COVID-19 treatments. Importantly, as age, race, gender, and comorbidities have been associated with COVID-19 risk [72], these factors are deemed warranted to assess their role in the variation of treatment responses and need further investigation.

Population genetics is also needed to better understand the association between genetic variability and COVID-19. The importance and complexity of population genetic information, such as genome-wide association studies (GWAS), for drug discovery are exemplified by a study showing that 8% of drugs approved by the FDA target molecules with genetic support, whereas only 2% of phase-1 drugs are genetically supported [73]. Despite such low rates, GWAS can help identify therapeutics that can be repurposed to treat individuals affected by diseases that are mechanistically related to those for which the drugs were developed [74]. Insights from GWAS can also inform about better patient management and therapy, such as the case of variants in six genes on chromosome 3, namely *SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6*, and *XCR1*, which have been recently associated with severe COVID-19 cases with respiratory failure [75].

Understanding population genetic heterogeneity is crucial for vaccine design, in particular, as it concerns the individual variability of the major histocompatibility complex (MHC-I and MHC-II) proteins, encoded by the *HLA* gene, which present SARS-CoV-2 epitopes to the immune system. Such individual variability, coupled with the importance of cellular immunity in the severity of the response to the infection, makes the identification of actionable targets for COVID-19 vaccines a challenging endeavor. AI models for COVID-19 vaccine development focus on the prediction of potential epitopes by using a variety of techniques, such as deep docking [76], long short-term memory networks [77], extreme gradient boosting [78], as well as approaches that account for different *HLA* alleles by combining several existing machine learning tools [79]. A recent survey of AI-based approaches to COVID-19 vaccine design [80] suggests that the most popular candidate is the SARS-CoV-2 spike protein, which initiates the interaction with the host through the attachment to the ACE2 receptor [81].

Mining of the Medical Literature

The staggering rate of publications about COVID-19, both in the form of preprints and peer-reviewed articles, is posing unprecedented challenges to knowledge acquisition and the information quality assessment process. A large part of content is produced by humans for humans, in the form of free text, where crucial pieces of information end up being buried. Because free text is not intelligible by machines, human intervention must identify the relevant pieces of information from the publications and turn it into a tabular form. Recent developments in NLP techniques have helped the automation of this process through machine learning and, in particular, deep learning algorithms [82,83]. Symptoms, patient demographics, clinical data, algorithms, performance, and limitations are identifiable in the texts by properly trained models, which can obtain comparable accuracy to humans at a much faster rate, making it finally possible to monitor the enormous volume of the literature produced [84]. The resulting structured data can be exploited to enrich knowledge graphs (KGs) [85-87], which provide a means to represent and formalize information [85,88], analytical, relational, and inferential investigations and fill the knowledge gaps in the community. Moreover, to rationalize the immense quantity of information on COVID-19, new algorithms

can generate low-dimensional representations of the KGs, allowing researchers for clustering and classification [85,89].

We list here representative KG efforts that have been directed at the fight against COVID-19 (see Textbox 1).

Textbox 1. Knowledge graph resources for COVID-19.

Project names and references:

- KG-Covid-19 Knowledge Graph Hub [90]
- COVID-19 Community Project [91]
- COVID-KG [92]
- CovidGraph [93]
- COVID-19 Miner [94]
- COVID-19 Biomedical Knowledge Miner [95]
- COVID-19 Taxila [96]

The KG-Covid-19 Knowledge Graph Hub project is the first Knowledge Graph Hub (KG-Hub) [90] dedicated to COVID-19. KG-Hub is a software to download and transform data to a central location for building KGs from different combinations of data sources. The Covid-19 KG-Hub downloads and transforms data from more than 50 different COVID-19 databases of drugs, genes, proteins, ontologies, diseases, phenotypes, and publications and generates a KG that can be used for machine learning.

The COVID-19 Community Project [91] is a community-based KG that links heterogeneous datasets about COVID-19, in three main areas: the host, the virus, and the cellular environment. These KGs use several publicly available datasets, such as the CORD-19 dataset, a set of over 51,000 scholarly articles about coronaviruses [97].

Other notable databases used in KGs are the COVID-19 Data Portal (see Introduction) and The COVID-19 Drug and Gene Set Library [98]. One of the tools that use these is the COVID-KG [92], which embeds entities in the KG, such as papers, authors, or journals [99].

CovidGraph [93] is a collaboration of researchers to build a research and communication platform that encompasses over 40,000 publications, case statistics, genes and functions, molecular data, and much more. The output is a KG in which entity relationships can be found and new pieces of literature can be discovered. Another tool that uses the CORD-19 dataset is COVID-19 Miner [94], which provides access to a database of interactions among genes or proteins, chemicals, and biological processes related to SARS-CoV-2, which are automatically extracted using NLP from the CORD-19 dataset and manuscripts updated daily from the preprint servers medRxiv and bioRxiv [100].

Furthermore, COVID-19 Biomedical Knowledge Miner [85,95] is an intent to lay the foundation for a comprehensive and interactive KG in the context of COVID-19 that connects the causes and effects and enables users to completely explore the information contained therein. Data are supplied from papers available in PubMed and preprints available from platforms such as bioRxiv, chemRxiv, medRxiv, PrePrints, and Research Square. Lastly, COVID-19 Taxila [96] is an AI and NLP system that uses thousands of COVID-19–related publications, clinical trials, and other relevant sources to enable users to search and analyze the COVID-19 literature. Publications and data are automatically updated.

Discussion

The COVID-19 pandemic has caused some of the most significant challenges that national health care systems have had to face in recent human history. These systems include not only hospitals but also a multitude of clinicians, retirement and nursing homes, families, and communities. Government lockdown policies undertaken to reduce hospital strain has impacted the society as a whole and has also had social and economic consequences, which have been more severe for minorities and vulnerable groups [101]. Moreover, this pandemic is taking place in the age of social media and Web 2.0, which contain plenty of misinformation and fake news, and with no way for the average internet user to check the reliability of the sources. Nevertheless, the COVID-19 crisis has also shown the promise of technology in facilitating a better understanding of a complex disease and its impact on public health.

Here, we illustrated examples of how AI can advance research and clinical medicine and prepare governments for future similar crises. AI shows promise to deliver models for outbreak analytics and detection, prevention, early intervention, and decision-making. We highlighted the unparalleled opportunity for AI to fill the gap between translational research and clinical medicine. Finally, in addition to the medical applications of AI, it is worth mentioning the potential of NLP for monitoring the quality of the information available to the public and fighting fake news [102-104].

Thanks to the availability of big data and high-performance computing, the fight against the novel coronavirus can leverage the support of AI, as demonstrated by initiatives such as the COVID-19 High Performance Computing Consortium [105]. This technology allows us to address, at a much higher speed and a comparable performance, complex tasks that cannot be executed by humans—who can now focus on more intelligence-demanding activities such as emotional intelligence and human-to-human bonding [106].

Although AI is traditionally trained on large datasets for identifying population-level patterns (ie, common characteristics among people belonging to some clinical classes), recent efforts have promoted the utilization of this technology in conjunction with the principles of precision medicine, to substitute the "average patient" [42] with a real individual, based on geographical and socioeconomic signature as well as genetic, epigenetic, and other molecular profiles [107]. Under this paradigm, AI is meant to empower clinicians to tailor interventions [108] (whether preventive or therapeutic) to the nuanced—and often unique—features of every human being [109]. To this end, multidimensional datasets, such as the variety of data modalities that are currently collected and modeled for COVID-19 [110-112], capture individual genetic, biochemical, physiological, environmental, and behavioral variations [113] that may interfere with the development, progression, and treatment of a disease. Thanks to the drop in price of sequencing the human genome (from billions to hundreds of dollars in 30 years [114]), it is now possible to exploit AI to study phenotypic, genotypic, and environmental correlations among diseases [115]. With this approach, AI can predict the risk of an individual to develop a disease and estimate the likelihood of success for a treatment. In the case of COVID-19, this could lead to a better allocation of resources and an improved match between treatments and patients, consequently improving outcomes for preventive and therapeutic interventions. Therefore, AI-aided precision medicine connects some of the key benefits for a sustainable and effective health care system: efficiency, efficacy, and safety assessment [30].

AI is recognized as a necessity to achieve precision medicine in COVID-19. The current crisis has highlighted that a huge amount of work is still needed to exploit AI-based solutions to their full potential in order to transform health care. AI implementation in the clinical setting is still far from completion [115]. The highly fragmented and diverse health care systems, absence of a protocol for documenting patient data, ethical constraints (such as privacy), and limitations of AI itself (eg, bias and non-interpretability) still represent serious challenges to extensive AI adoption [116].

Acknowledgments

We are deeply thankful to the Women's Brain Project (WBP) (www.womensbrainproject.com), an international organization advocating for women's brain and mental health through scientific research, debate, and public engagement. The authors would like to thank Maria Teresa Ferretti and Shahnaz Radjy for the helpful comments. DC, AM, and AV have received funding from the European Commission's Horizon 2020 Program H2020-SC1-DTH-2018-1, "iPC-individualizedPaediatricCure" (ref. 826121), and H2020-ICT-2018-2, "INFORE-Interactive Extreme-Scale Analytics and Forecasting" (ref. 825070).

Authors' Contributions

ES and NM conceived the study with the contribution of ASC, EC, and DC. ES and NM directed the content selection and design, assisted by EC, DC, and AM. AV, KH, and CL supervised the project. The corresponding author had the final responsibility for the decision to submit the manuscript for publication. All authors have contributed to the writing and editing of the manuscript and have read and approved the final manuscript.

Conflicts of Interest

ASC and ES are currently employees at Biogen International GmbH, HQ, Switzerland, and Bayer Pharmaceuticals, USA, respectively. The other authors declare no competing interests. KH is a founder and owns equity of CRA Health (formerly Hughes RiskApps), is co-creator of Ask2Me.Org, which is licensed for commercial use by the Dana-Farber Cancer Institute, and receives honoraria from Myriad Genetics.

References

1. Legido-Quigley H, Asgari N, Teo YY, Leung GM, Oshitani H, Fukuda K, et al. Are high-performing health systems resilient against the COVID-19 epidemic? The Lancet 2020 Mar;395(10227):848-850. [doi: 10.1016/S0140-6736(20)30551-1]
2. McCall B. COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread. The Lancet Digital Health 2020 Apr;2(4):e166-e167. [doi: 10.1016/s2589-7500(20)30054-6]
3. Niiler E. An AI Epidemiologist Sent the First Alerts of the Coronavirus Internet. Wired. 2020 Jan 25. URL: https://www.wired.com/story/ai-epidemiologist-wuhan-public-health-warnings/ [accessed 2020-07-07]
4. Computational predictions of protein structures associated with COVID-19 Internet. DeepMind. URL: https://deepmind.com/research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19 [accessed 2020-07-07]
5. Lan L, Xu D, Ye G, Xia C, Wang S, Li Y, et al. Positive RT-PCR test results in patients recovered from COVID-19. JAMA 2020 Apr 21;323(15):1502-1503 [FREE Full text] [doi: 10.1001/jama.2020.2783] [Medline: 32105304]
6. Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, COPE Consortium. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. Science 2020 Jun 19;368(6497):1362-1367 [FREE Full text] [doi: 10.1126/science.abc0473] [Medline: 32371477]
7. Shaban-Nejad A, Michalowski M, Buckeridge DL. Health intelligence: how artificial intelligence transforms population and personalized health. NPJ Digit Med 2018;1:53 [FREE Full text] [doi: 10.1038/s41746-018-0058-9] [Medline: 31304332]

8. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf 2019 Mar;28(3):231-237. [doi: 10.1136/bmjqs-2018-008370] [Medline: 30636200]
9. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. JAMA. Epub ahead of print posted online on November 22, 2019 2019. [doi: 10.1001/jama.2019.18058] [Medline: 31755905]
10. Nelson GS. Bias in artificial intelligence. N C Med J 2019;80(4):220-222 [FREE Full text] [doi: 10.18043/ncm.80.4.220] [Medline: 31278182]
11. Cirillo D, Catuara-Solarz S, Morey C, Guney E, Subirats L, Mellino S, et al. Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ Digit Med 2020 Jun 01;3(1):81 [FREE Full text] [doi: 10.1038/s41746-020-0288-5] [Medline: 33597695]
12. Larrazabal AJ, Nieto N, Peterson V, Milone DH, Ferrante E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. Proc Natl Acad Sci U S A 2020 Jun 09;117(23):12592-12594 [FREE Full text] [doi: 10.1073/pnas.1919012117] [Medline: 32457147]
13. The impact of sexgender in the COVID-19 pandemic Internet. European Commission. 2020. URL: https://ec.europa.eu/info/publications/impact-sex-and-gender-covid-19-pandemic_en [accessed 2020-07-07]
14. Shilo S, Rossman H, Segal E. Axes of a revolution: challenges and promises of big data in healthcare. Nat Med 2020 Jan;26(1):29-38. [doi: 10.1038/s41591-019-0727-5] [Medline: 31932803]
15. COVID-19 Data Portal. URL: https://www.covid19dataportal.org/ [accessed 2020-10-06]
16. National COVID Cohort Collaborative (N3C) Internet. National Center for Advancing Translational Sciences. URL: https://ncats.nih.gov/n3c [accessed 2020-10-06]
17. The COVID-19 Host Genetics Initiative. COVID-19 hg. URL: https://www.covid19hg.org/ [accessed 2020-10-06]
18. COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. Eur J Hum Genet 2020 Jun;28(6):715-718 [FREE Full text] [doi: 10.1038/s41431-020-0636-6] [Medline: 32404885]
19. The COVID-19 Sex-Disaggregated Data Tracker. Global Health 50/50 - The Sex, Gender, and COVID-19 Project. URL: https://globalhealth5050.org/the-sex-gender-and-covid-19-project/ [accessed 2020-10-06]
20. Lindvall C, Cassel CK, Pantilat SZ, DeCamp M. Ethical considerations in the use of AI mortality predictions in the care of people with serious illness. Health Affairs. URL: https://www.healthaffairs.org/do/10.1377/hblog20200911.401376/full/ [accessed 2020-10-04]
21. Progress using COVID-19 patient data to train machine learning models for healthcare. University of Cambridge. URL: https://www.cam.ac.uk/research/news/progress-using-covid-19-patient-data-to-train-machine-learning-models-for-healthcare [accessed 2020-07-07]
22. Rahmatizadeh S, Valizadeh-Haghi S, Dabbagh A. The role of artificial intelligence in management of critical COVID-19 patients. Journal of Cellular & Molecular Anesthesia 2020 Apr;5(1):16-22. [doi: 10.22037/jcma.v5i1.29752]
23. Grasselli G, Pesenti A, Cecconi M. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: Early experience and forecast during an emergency response. JAMA 2020 Apr 28;323(16):1545-1546. [doi: 10.1001/jama.2020.4031] [Medline: 32167538]
24. Dawson NV, Arkes HR. Systematic errors in medical decision making: judgment limitations. J Gen Intern Med 1987;2(3):183-187. [doi: 10.1007/BF02596149] [Medline: 3295150]
25. Sendak MP, Ratliff W, Sarro D, Alderton E, Futoma J, Gao M, et al. Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study. JMIR Med Inform 2020 Jul 15;8(7):e15182 [FREE Full text] [doi: 10.2196/15182] [Medline: 32673244]
26. Verghese A, Shah NH, Harrington RA. What this computer needs is a physician: humanism and artificial intelligence. JAMA 2018 Jan 02;319(1):19-20. [doi: 10.1001/jama.2017.19198] [Medline: 29261830]
27. Patel BN, Rosenberg L, Willcox G, Baltaxe D, Lyons M, Irvin J, et al. Human-machine partnership with artificial intelligence for chest radiograph diagnosis. NPJ Digit Med 2019;2:111 [FREE Full text] [doi: 10.1038/s41746-019-0189-7] [Medline: 31754637]
28. Vindegaard N, Benros ME. COVID-19 pandemic and mental health consequences: systematic review of the current evidence. Brain Behav Immun 2020 Oct;89:531-542 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.048] [Medline: 32485289]
29. Figueroa CA, Aguilera A. The need for a mental health technology revolution in the COVID-19 pandemic. Front Psychiatry 2020;11:523 [FREE Full text] [doi: 10.3389/fpsyt.2020.00523] [Medline: 32581891]
30. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
31. Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. NPJ Digit Med 2019;2:77 [FREE Full text] [doi: 10.1038/s41746-019-0155-4] [Medline: 31453372]
32. Lin SY, Mahoney MR, Sinsky CA. Ten ways artificial intelligence will transform primary care. J Gen Intern Med 2019 Aug;34(8):1626-1630 [FREE Full text] [doi: 10.1007/s11606-019-05035-1] [Medline: 31090027]
33. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AICONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med 2020 Sep;26(9):1364-1374 [FREE Full text] [doi: 10.1038/s41591-020-1034-x] [Medline: 32908283]

34. Cruz Rivera S, Liu X, Chan A, Denniston A, Calvert M, SPIRIT-AICONSORT-AI Working Group, SPIRIT-AICONSORT-AI Steering Group, SPIRIT-AICONSORT-AI Consensus Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med 2020 Sep;26(9):1351-1363 [FREE Full text] [doi: 10.1038/s41591-020-1037-7] [Medline: 32908284]
35. Filipp FV. Opportunities for artificial intelligence in advancing precision medicine. Curr Genet Med Rep 2019 Dec;7(4):208-213 [FREE Full text] [doi: 10.1007/s40142-019-00177-4] [Medline: 31871830]
36. Zeggini E, Gloyn AL, Barton AC, Wain LV. Translational genomics and precision medicine: moving from the lab to the clinic. Science 2019 Sep 27;365(6460):1409-1413. [doi: 10.1126/science.aax4588] [Medline: 31604268]
37. Gruson D, Helleputte T, Rousseau P, Gruson D. Data science, artificial intelligence, and machine learning: opportunities for laboratory medicine and the value of positive regulation. Clin Biochem 2019 Jul;69:1-7. [doi: 10.1016/j.clinbiochem.2019.04.013] [Medline: 31022391]
38. Suspected COVID-19 pneumonia Diagnosis Aid System. URL: https://intensivecare.shinyapps.io/COVID19/ [accessed 2021-02-19]
39. Feng C, Huang Z, Wang L, Chen X, Zhai Y, Zhu F, et al. A novel triage tool of artificial intelligence assisted diagnosis aid system for suspected COVID-19 pneumonia in fever clinics. medRxiv. Preprint posted online on March 20, 2020. [doi: 10.1101/2020.03.19.20039099]
40. Jiang X, Coffee M, Bari A, Wang J, Jiang X, Huang J, et al. Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. Computers, Materials & Continua 2020 Mar;63(1):537-551 [FREE Full text] [doi: 10.32604/cmc.2020.010691]
41. Freedman AN, Seminara D, Gail MH, Hartge P, Colditz GA, Ballard-Barbash R, et al. Cancer risk prediction models: a workshop on development, evaluation, and application. J Natl Cancer Inst 2005 May 18;97(10):715-723. [doi: 10.1093/jnci/dji128] [Medline: 15900041]
42. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A deep learning mammography-based model for improved breast cancer risk prediction. Radiology 2019 Jul;292(1):60-66. [doi: 10.1148/radiol.2019182716] [Medline: 31063083]
43. Hu S, Santus E, Forsyth AW, Malhotra D, Haimson J, Chatterjee NA, et al. Can machine learning improve patient selection for cardiac resynchronization therapy? PLoS One 2019;14(10):e0222397 [FREE Full text] [doi: 10.1371/journal.pone.0222397] [Medline: 31581234]
44. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]
45. Mei X, Lee H, Diao K, Huang M, Lin B, Liu C, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. Nat Med 2020 Aug;26(8):1224-1228 [FREE Full text] [doi: 10.1038/s41591-020-0931-3] [Medline: 32427924]
46. Jin C, Chen W, Cao Y, Xu Z, Zhang X, Deng L, et al. Development and Evaluation of an AI System for COVID-19 Diagnosis. medRxiv. Preprint posted online on June 02, 2020. [doi: 10.1101/2020.03.20.20039834]
47. Schaffter T, Buist DSM, Lee CI, Nikulin Y, Ribli D, Guan Y, the DM DREAM Consortium, et al. Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms. JAMA Netw Open 2020 Mar 02;3(3):e200265 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.0265] [Medline: 32119094]
48. Xiao L, Li P, Sun F, Zhang Y, Xu C, Zhu H, et al. Development and validation of a deep learning-based model using computed tomography imaging for predicting disease severity of coronavirus disease 2019. Front Bioeng Biotechnol 2020;8:898 [FREE Full text] [doi: 10.3389/fbioe.2020.00898] [Medline: 32850746]
49. Clinical management of severe acute respiratory infection (SARI) when COVID-19 disease is suspected. World Health Organization. 2020 Mar 13. URL: https://www.who.int/docs/default-source/coronaviruse/clinical-management-of-novel-cov.pdf?sfvrsn=bc7da517_2 [accessed 2021-02-19]
50. NIH harnesses AI for COVID-19 diagnosis, treatment, and monitoring. National Institutes of Health. 2020 Aug 05. URL: https://www.nih.gov/news-events/news-releases/nih-harnesses-ai-covid-19-diagnosis-treatment-monitoring [accessed 2020-10-06]
51. Yan L, Zhang H, Goncalves J, Xiao Y, Wang M, Guo Y, et al. A machine learning-based model for survival prediction in patients with severe COVID-19 infection. medRxiv. Preprint posted online on March 17, 2020. [FREE Full text] [doi: 10.1101/2020.02.27.20028027]
52. Senanayake SL. Drug repurposing strategies for COVID-19. Future Drug Discovery 2020 Apr 01;2(2) [FREE Full text] [doi: 10.4155/fdd-2020-0010]
53. Zhavoronkov A, Aladinskiy V, Zhebrak A, Zagribelnyy B, Terentiev V, Bezrukov D, et al. Potential COVID-2019 3C-like Protease Inhibitors Designed Using Generative Deep Learning Approaches. ChemRxiv. Preprint posted online on February 11, 2020. [FREE Full text]
54. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 2020 Jul;583(7816):459-468 [FREE Full text] [doi: 10.1038/s41586-020-2286-9] [Medline: 32353859]
55. Richardson P, Griffin I, Tucker C, Smith D, Oechsle O, Phelan A, et al. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. The Lancet 2020 Feb;395(10223):e30-e31. [doi: 10.1016/s0140-6736(20)30304-4]

56. Beck BR, Shin B, Choi Y, Park S, Kang K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. Comput Struct Biotechnol J 2020;18:784-790 [FREE Full text] [doi: 10.1016/j.csbj.2020.03.025] [Medline: 32280433]
57. Ostaszewski M, Mazein A, Gillespie ME, Kuperstein I, Niarakis A, Hermjakob H, et al. COVID-19 disease map, building a computational repository of SARS-CoV-2 virus-host interaction mechanisms. Sci Data 2020 May 05;7(1):136 [FREE Full text] [doi: 10.1038/s41597-020-0477-8] [Medline: 32371892]
58. Gysi D, Valle ID, Zitnik M, Ameli A, Gan X, Varol O, et al. Network medicine framework for identifying drug repurposing opportunities for COVID-19. arXiv. Preprint posted online on April 15, 2020. [FREE Full text]
59. Fleming N. How artificial intelligence is changing drug discovery. Nature 2018 May;557(7707):S55-S57. [doi: 10.1038/d41586-018-05267-x] [Medline: 29849160]
60. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. Cell 2020 Feb 20;180(4):688-702.e13. [doi: 10.1016/j.cell.2020.01.021] [Medline: 32084340]
61. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. Nature 2020 Jan;577(7792):706-710. [doi: 10.1038/s41586-019-1923-7] [Medline: 31942072]
62. Nelson MR, Johnson T, Warren L, Hughes AR, Chissoe SL, Xu C, et al. The genetics of drug efficacy: opportunities and challenges. Nat Rev Genet 2016 Apr;17(4):197-206. [doi: 10.1038/nrg.2016.12] [Medline: 26972588]
63. Tan-Koi WC, Leow PC, Teo YY. Applications of pharmacogenomics in regulatory science: a product life cycle review. Pharmacogenomics J 2018 May 22;18(3):359-366. [doi: 10.1038/tpj.2017.47] [Medline: 29205206]
64. Weinshilboum RM, Wang L. Pharmacogenomics: precision medicine and drug response. Mayo Clin Proc 2017 Nov;92(11):1711-1722 [FREE Full text] [doi: 10.1016/j.mayocp.2017.09.001] [Medline: 29101939]
65. Lauschke VM, Ingelman-Sundberg M. Emerging strategies to bridge the gap between pharmacogenomic research and its clinical implementation. NPJ Genom Med 2020 Mar 05;5(1):9 [FREE Full text] [doi: 10.1038/s41525-020-0119-2] [Medline: 33579994]
66. Silverman EK. Genetics of COPD. Annu Rev Physiol 2020 Feb 10;82:413-431 [FREE Full text] [doi: 10.1146/annurev-physiol-021317-121224] [Medline: 31730394]
67. Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome. NCBI. URL: https://www.ncbi.nlm.nih.gov/nuccore/MN908947 [accessed 2020-10-04]
68. Real-time tracking of pathogen evolution - SARS-CoV-2 (COVID-19). Nextstrain. URL: https://nextstrain.org/ [accessed 2021-02-22]
69. Kachuri L, Francis S, Morrison M, Wendt GA, Bossé Y, Cavazos TB, et al. The landscape of host genetic factors involved in immune response to common viral infections. Genome Med 2020 Oct 27;12(1):93 [FREE Full text] [doi: 10.1186/s13073-020-00790-x] [Medline: 33109261]
70. Benetti E, Tita R, Spiga O, Ciolfi A, Birolo G, Bruselles A, GEN-COVID Multicenter Study, et al. ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. Eur J Hum Genet 2020 Nov;28(11):1602-1614 [FREE Full text] [doi: 10.1038/s41431-020-0691-z] [Medline: 32681121]
71. Takahashi T, Luzum JA, Nicol MR, Jacobson PA. Pharmacogenomics of COVID-19 therapies. NPJ Genom Med 2020;5:35 [FREE Full text] [doi: 10.1038/s41525-020-00143-y] [Medline: 32864162]
72. Wenham C, Smith J, Morgan R. COVID-19: the gendered impacts of the outbreak. The Lancet 2020 Mar;395(10227):846-848. [doi: 10.1016/s0140-6736(20)30526-2]
73. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. Nat Genet 2015 Aug;47(8):856-860. [doi: 10.1038/ng.3314] [Medline: 26121088]
74. Dugger SA, Platt A, Goldstein DB. Drug development in the era of precision medicine. Nat Rev Drug Discov 2018 Mar;17(3):183-196 [FREE Full text] [doi: 10.1038/nrd.2017.226] [Medline: 29217837]
75. Severe Covid-19 GWAS Group, Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, et al. Genomewide association study of severe Covid-19 with respiratory failure. N Engl J Med 2020 Oct 15;383(16):1522-1534 [FREE Full text] [doi: 10.1056/NEJMoa2020283] [Medline: 32558485]
76. Ton A, Gentile F, Hsing M, Ban F, Cherkasov A. Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. Mol Inform 2020 Aug;39(8):e2000028 [FREE Full text] [doi: 10.1002/minf.202000028] [Medline: 32162456]
77. Abbasi B, Saraf D, Sharma T, Sinha R, Singh S, Gupta P, et al. OSF Preprints. Preprint posted online April 08, 2020. [doi: 10.31219/osf.io/f8zyw]
78. Ong E, Wang H, Wong MU, Seetharaman M, Valdez N, He Y. Vaxign-ML: supervised machine learning reverse vaccinology model for improved prediction of bacterial protective antigens. Bioinformatics 2020 May 01;36(10):3185-3191. [doi: 10.1093/bioinformatics/btaa119] [Medline: 32096826]
79. Malone B, Simovski B, Moliné C, Cheng J, Gheorghe M, Fontenelle H, et al. Artificial intelligence predicts the immunogenic landscape of SARS-CoV-2: toward universal blueprints for vaccine designs. bioRxiv. Preprint posted online on April 21, 2020. [FREE Full text] [doi: 10.1101/2020.04.21.052084]
80. Keshavarzi Arshadi A, Webb J, Salem M, Cruz E, Calad-Thomson S, Ghadirian N, et al. Artificial intelligence for COVID-19 drug discovery and vaccine development. Front Artif Intell 2020 Aug 18;3. [doi: 10.3389/frai.2020.00065]

81. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 2020 Apr 16;181(2):271-280.e8 [FREE Full text] [doi: 10.1016/j.cell.2020.02.052] [Medline: 32142651]
82. Santus E, Li C, Yala A, Peck D, Soomro R, Faridi N, et al. Do neural information extraction algorithms generalize across institutions? JCO Clin Cancer Inform 2019 Jul;3:1-8 [FREE Full text] [doi: 10.1200/CCI.18.00160] [Medline: 31310566]
83. Santus E, Schuster T, Tahmasebi AM, Li C, Yala A, Lanahan CR, et al. Exploiting rules to enhance machine learning in extracting information from multi-institutional prostate pathology reports. JCO Clin Cancer Inform 2020 Oct;4:865-874 [FREE Full text] [doi: 10.1200/CCI.20.00028] [Medline: 33006906]
84. Adnan K, Akbar R. An analytical study of information extraction from unstructured and multidimensional big data. J Big Data 2019 Oct 17;6:91. [doi: 10.1186/s40537-019-0254-8]
85. Domingo-Fernández D, Baksi S, Schultz B, Gadiya Y, Karki R, Raschka T, et al. COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology. bioRxiv. Preprint posted online on April 15, 2020. [doi: 10.1101/2020.04.14.040667]
86. Cowen L, Ideker T, Raphael BJ, Sharan R. Network propagation: a universal amplifier of genetic associations. Nat Rev Genet 2017 Sep;18(9):551-562. [doi: 10.1038/nrg.2017.38] [Medline: 28607512]
87. Humayun F, Domingo-Fernández D, Paul George AA, Hopp M, Syllwasschy BF, Detzel MS, et al. A computational approach for mapping heme biology in the context of hemolytic disorders. Front Bioeng Biotechnol 2020;8:74 [FREE Full text] [doi: 10.3389/fbioe.2020.00074] [Medline: 32211383]
88. Nelson W, Zitnik M, Wang B, Leskovec J, Goldenberg A, Sharan R. To embed or not: network embedding as a paradigm in computational biology. Front Genet 2019;10:381 [FREE Full text] [doi: 10.3389/fgene.2019.00381] [Medline: 31118945]
89. Hamilton WL, Ying R, Leskovec J. Representation learning on graphs: methods and applications. IEEE Data Engineering Bulletin 2017;40(3):52-74 [FREE Full text]
90. KG-Covid-19 Knowledge Graph Hub. GitHub. URL: https://github.com/Knowledge-Graph-Hub/kg-covid-19 [accessed 2021-02-22]
91. COVID-19 Community Project. GitHub. URL: https://github.com/covid-19-net/covid-19-community [accessed 2021-02-22]
92. COVID-KG. GitHub. URL: https://github.com/GillesVandewiele/COVID-KG/ [accessed 2021-02-22]
93. COVID-19 Knowledge Graph. CovidGraph. URL: https://covidgraph.org/ [accessed 2021-02-22]
94. COVID-19 Miner. GitHub. URL: https://github.com/rupertoverall/covid19-miner [accessed 2021-02-22]
95. COVID-19 Biomedical Knowledge Miner. Biomedical Knowledge Miner (BiK>Mi). URL: https://bikmi.covid19-knowledgespace.de/ [accessed 2021-02-22]
96. COVID-19 Taxila. URL: http://covid19.taxila.io/ [accessed 2021-02-22]
97. COVID-19 Open Research Dataset Challenge (CORD-19). Kaggle (Allen Institute For AI). URL: https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge [accessed 2020-10-04]
98. The COVID-19 Drug and Gene Set Library. Ma'ayan Laboratory, Computational Systems Biology. URL: https://amp.pharm.mssm.edu/covid19/ [accessed 2021-02-22]
99. Ristoski P, Rosati J, Di Noia T, De Leone R, Paulheim H. RDF2Vec: RDF graph embeddings and their applications. SW 2019 May 23;10(4):721-752. [doi: 10.3233/sw-180317]
100. COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv. URL: https://connect.biorxiv.org/relate/content/181 [accessed 2020-10-04]
101. The Lancet Public Health. COVID-19 puts societies to the test. The Lancet Public Health 2020 May;5(5):e235. [doi: 10.1016/s2468-2667(20)30097-9]
102. Alam F, Dalvi F, Shaar S, Durrani N, Mubarak H, Nikolov A, et al. Fighting the COVID-19 infodemic in social media: a holistic perspective and a call to arms. arXiv. Preprint posted online on July 15, 2020 [FREE Full text]
103. Song X, Petrak J, Jiang Y, Singh I, Maynard D, Bontcheva K. Classification aware neural topic model and its application on a new COVID-19 disinformation corpus. arXiv. Preprint posted online on June 5, 2020 [FREE Full text]
104. Li Y, Grandison T, Silveyra P, Douraghy A, Guan X, Kieselbach T, et al. Jennifer for COVID-19: An NLP-powered chatbot built for the people and by the people to combat misinformation. : Association for Computational Linguistics; 2020 Presented at: Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020; July 2020; Online URL: https://www.aclweb.org/anthology/2020.nlpcovid19-acl.9
105. The COVID-19 High Performance Computing Consortium. URL: https://covid19-hpc-consortium.org/ [accessed 2021-02-22]
106. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. NPJ Digit Med 2018;1:5 [FREE Full text] [doi: 10.1038/s41746-017-0012-2] [Medline: 31304291]
107. Shen B, Yi X, Sun Y, Bi X, Du J, Zhang C, et al. Proteomic and metabolomic characterization of COVID-19 patient sera. Cell 2020 Jul 09;182(1):59-72.e15 [FREE Full text] [doi: 10.1016/j.cell.2020.05.032] [Medline: 32492406]
108. Ahmed Z, Mohamed K, Zeeshan S, Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database (Oxford) 2020 Jan 01;2020:baaa010 [FREE Full text] [doi: 10.1093/database/baaa010] [Medline: 32185396]

109. Shah P, Kendall F, Khozin S, Goosen R, Hu J, Laramie J, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. NPJ Digit Med 2019;2:69 [FREE Full text] [doi: 10.1038/s41746-019-0148-3] [Medline: 31372505]
110. Mahalle P, Sable N, Mahalle N, Shinde G. Data analytics: COVID-19 prediction using multimodal data. Preprints. Preprint published online on April 16, 2020. [doi: 10.20944/preprints202004.0257.v2]
111. Overmyer KA, Shishkova E, Miller IJ, Balnis J, Bernstein MN, Peters-Clarke TM, et al. Large-scale multi-omic analysis of COVID-19 severity. medRxiv. Preprint posted online on July 19, 2020 [FREE Full text] [doi: 10.1101/2020.07.17.20156513] [Medline: 32743614]
112. Ray S, Srivastava S. COVID-19 pandemic: hopes from proteomics and multiomics research. OMICS 2020 Aug;24(8):457-459. [doi: 10.1089/omi.2020.0073] [Medline: 32427517]
113. Schork N. Artificial intelligence and personalized medicine. In: Von Hoff D, Han H, editors. Precision Medicine in Cancer Therapy. Cancer Treatment and Research. Cham: Springer; 2019:265-283.
114. The Cost of Sequencing a Human Genome. National Human Genome Research Institute. URL: https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost [accessed 2020-10-06]
115. Dias R, Torkamani A. Artificial intelligence in clinical and genomic diagnostics. Genome Med 2019 Nov 19;11(1):70 [FREE Full text] [doi: 10.1186/s13073-019-0689-8] [Medline: 31744524]
116. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019 Oct 29;17(1):195. [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]

Abbreviations

ACE2: angiotensin-converting enzyme 2
AI: artificial intelligence
AUROC: area under the receiver operating characteristic curve
CNN: convolutional neural network
CONSORT-AI: Consolidated Standards of Reporting Trials–Artificial Intelligence
CT: computed tomography
FDA: Food and Drug Administration
GWAS: genome-wide association studies
HLA: human leukocyte antigen
ICU: intensive care unit
KG: knowledge graph
LASSO: least absolute shrinkage and selection operator
NLP: natural language processing
SPIRIT-AI: Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence

Edited by G Eysenbach, C Basch; submitted 12.07.20; peer-reviewed by K Ferrini, M Afzal; comments to author 17.08.20; revised version received 07.10.20; accepted 31.01.21; published 12.03.21.

Please cite as:
Santus E, Marino N, Cirillo D, Chersoni E, Montagud A, Santuccione Chadha A, Valencia A, Hughes K, Lindvall C
Artificial Intelligence–Aided Precision Medicine for COVID-19: Strategic Areas of Research and Development
J Med Internet Res 2021;23(3):e22453
URL: https://www.jmir.org/2021/3/e22453
doi:10.2196/22453
PMID:33560998

Original Paper

COVID-19 Discourse on Twitter in Four Asian Countries: Case Study of Risk Communication

Sungkyu Park[1*], PhD; Sungwon Han[2*], MSc; Jeongwook Kim[2*], MSc; Mir Majid Molaie[2], BSc; Hoang Dieu Vu[3], MSc; Karandeep Singh[1], PhD; Jiyoung Han[2], PhD; Wonjae Lee[2], PhD; Meeyoung Cha[1,2], PhD

[1]Data Science Group, Institute for Basic Science, Daejeon, Republic of Korea

[2]Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

[3]Electrical and Electronic Engineering, Phenikaa University, Hanoi, Vietnam

[*]these authors contributed equally

Corresponding Author:
Meeyoung Cha, PhD
Data Science Group
Institute for Basic Science
55, Expo-ro, Yuseong-gu
Daejeon, 34126
Republic of Korea
Phone: 82 428788114
Fax: 82 428788079
Email: meeyoung.cha@gmail.com

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/3/e28926

Abstract

Background: COVID-19, caused by SARS-CoV-2, has led to a global pandemic. The World Health Organization has also declared an infodemic (ie, a plethora of information regarding COVID-19 containing both false and accurate information circulated on the internet). Hence, it has become critical to test the veracity of information shared online and analyze the evolution of discussed topics among citizens related to the pandemic.

Objective: This research analyzes the public discourse on COVID-19. It characterizes risk communication patterns in four Asian countries with outbreaks at varying degrees of severity: South Korea, Iran, Vietnam, and India.

Methods: We collected tweets on COVID-19 from four Asian countries in the early phase of the disease outbreak from January to March 2020. The data set was collected by relevant keywords in each language, as suggested by locals. We present a method to automatically extract a time–topic cohesive relationship in an unsupervised fashion based on natural language processing. The extracted topics were evaluated qualitatively based on their semantic meanings.

Results: This research found that each government's official phases of the epidemic were not well aligned with the degree of public attention represented by the daily tweet counts. Inspired by the issue-attention cycle theory, the presented natural language processing model can identify meaningful transition phases in the discussed topics among citizens. The analysis revealed an inverse relationship between the tweet count and topic diversity.

Conclusions: This paper compares similarities and differences of pandemic-related social media discourse in Asian countries. We observed multiple prominent peaks in the daily tweet counts across all countries, indicating multiple issue-attention cycles. Our analysis identified which topics the public concentrated on; some of these topics were related to misinformation and hate speech. These findings and the ability to quickly identify key topics can empower global efforts to fight against an infodemic during a pandemic.

(J Med Internet Res 2021;23(3):e23272) doi:10.2196/23272

KEYWORDS

COVID-19; coronavirus; infodemic; infodemiology; infoveillance; Twitter; topic phase detection; topic modeling; latent Dirichlet allocation; risk communication

Introduction

Background

The COVID-19 pandemic has affected global health and the economy. The use of social media and the internet to seek and share information about the virus has increased rapidly [1,2], which makes them excellent media to examine for patterns of risk communication during a pandemic. During this time, one could observe how the intentional and unintentional spread of misinformation (here defined as unconfirmed or false information) jeopardized public health on such platforms. Studies have shown that people tend to share misinformation faster and more widely than real information [3-5]. The sheer amount of data and the mixture of accurate and false information leaves people confused over which safety guidelines and health tips to follow. This phenomenon has been called an *infodemic* [6]. Infodemics have become a real threat; misinformation on COVID-19 has shifted from focusing on false preventive measures to antivaccination arguments [7] and vandalism toward telecommunication infrastructures [8].

Analysis of risk communication is critical because it helps better understand how and why people propagate or consume certain information upon a threat to their health, economic, or social well-being. Such analysis helps stakeholders prepare and reach informed conclusions about how their decisions affect individuals' interests, values, and well-being [9]. In the context of COVID-19, which is our interest, analysis of risk communication can find opportunities to mitigate the propagation of false claims that threaten public safety [10].

Studies have identified online risk communication topics by collectively considering temporal tweet trends by adopting, for instance, a statistical clustering method that scans over time [11,12] or a deep learning–based embedding and clustering method [13]. One limitation of statistical approaches is that inaccurate or incomplete input data can act as noise, resulting in unstable clustering results [14]. Embedding approaches for topic modeling have also required that one specify the time duration (eg, monthly). However, such an arbitrary division hinders finding natural topical transitions and critical risk communication topics. Therefore, flexible time durations are preferable in identifying topical shifts.

This research used the data gathered from social media to understand public discourse on COVID-19. Understanding public concerns will help determine which unproven claims or pieces of misinformation need to be debunked first and will contribute to fighting the disease. Primarily, we aim to identify what people say without gatekeeping. For instance, identifying new misinformation in countries that are experiencing a pandemic at an early stage can buy time to debunk the same piece of misinformation in other countries before it poses a threat to public health [10].

To detect meaningful topical shifts of risk communication, one needs to demarcate temporal phases from the public discourse that reflect prevailing circumstances in the real world. If social media conversations were to change by the epidemic phases announced by local governments, one might use the same phases. However, government announcements do not necessarily match with the public interest. Following the issue-attention cycle theory [15], we leveraged drastic changes in the daily tweet volume to divide COVID-19 public discourse online in finding temporal phases. We extracted topics corresponding to the preset temporal phases based on a natural language processing method.

We used a spatiotemporal approach and considered tweets from different countries to provide more holistic views of risk communication. We present views from four Asian countries. Such a multicountry view was used to explore possible opportunities for joint efforts in managing risk communication. For example, early detection of misinformation can help social media services, social media communicators, journalists, policy makers, and medical professionals fight infodemics worldwide.

We ask the following research questions (RQs):

- RQ1: Do the official epidemic phases announced by governments reflect online interaction patterns?
- RQ2: Can topic phases be demarcated automatically based on a bottom-up approach?
- RQ3: What are the major topics corresponding to each topic phase?
- RQ4: What are the unique traits of the topic trends by country, and what are the distinguishing online communicative characteristics?

By answering these RQs, this study makes four contributions. First, we propose an end-to-end method of extracting risk communication topics in a spatial–temporal fashion with less gatekeeping. Second, we provide a theoretical ground (issue-attention cycle) to the framework and successfully assess its validity by observing multiple prominent peaks in the daily conversation. Third, we demonstrate via a case study of four countries a common risk communication trait. During the peak moments of conversation, users on social media concentrate on a few topics. Finally, we show from the case study which topics were directly linked to misinformation and hateful speech in the studied data.

The gathered data from Twitter and the codes (including language tokenizers and analysis codes) are accessible in Multimedia Appendix 1 and on GitHub [16].

Related Research

Issue-Attention Cycle

The issue-attention cycle model [15] conceptualizes how an issue rises into and fades away from the center of public attention. In the first stage, labeled the *preproblem* stage, an undesirable social condition (eg, the appearance of COVID-19)

emerges but does not yet draw much public attention. The second stage, dubbed *alarmed discovery and euphoric enthusiasm*, occurs when a triggering event (eg, the national spike of newly confirmed cases of COVID-19) heightens public awareness of the issue. In the third stage, called *realizing the cost of significant progress*, people begin to recognize the hardship involved in restructuring society, and individuals must sacrifice to solve the problem. This causes a *gradual decline of intense public interest*, the fourth stage. In the final *postproblem* stage, the current issue is replaced by a new issue and moves into a twilight zone of reduced public attention.

Not all issues follow the five stages of the issue-attention cycle [17]. As the cyclical patterns of public attention evolve, a wide array of public discourse has been found across multiple issues of climate change [18], emerging technologies [19,20], and public health risks [21,22]. There are also cultural differences in such discourse patterns. For example, concerning the H1N1 pandemic, South Korean news coverage showed five phases of increasing or decreasing attention. The corresponding US news coverage of the pandemic saw only two phases during the same 7-month time period [23].

Despite these fragmented findings, the issue-attention cycle framework provides insights into how public attention dramatically waxes and wanes. An issue that has gone through the cycle is different from issues that have not gone through the cycle in at least two ways. First, when an issue has achieved national prominence, new institutions, programs, and measures will have been developed to address the situation. These developments and their societal impacts are likely to persist even after public attention has shifted elsewhere. Second, the prolonged impacts of these developments are shaped by what was heavily discussed when the issue was of primary public concern.

Although the issue-attention cycle was initially proposed to model traditional media such as newspapers and television, there is a burgeoning literature applying the model to social media platforms. Among them, Twitter serves as a forum that the public is increasingly turning toward to seek and share information that is not subjected to a gatekeeping process [24]. It has become common for journalists to refer to tweets in their news stories. Research has also found that Twitter takes the lead in and exerts control over public discourse, particularly in the early stages of an issue-attention cycle [20,25].

Building on these prior studies, we analyzed Twitter conversations about COVID-19 to examine social media's issue-attention cycle. We present how to build an end-to-end method of identifying meaningful *topic phases* dynamically. This allows us to compare how issue-attention cycles appear in different countries on the same catastrophic event. To the best of our knowledge, no study has applied dynamic topic modeling in the context of risk communication.

COVID-19–Related Analyses

Studies have examined various impacts of the pandemic. Researchers have focused on predicting the transmissibility of the virus. One study estimated the viral reproduction number (R_0) of SARS-CoV-2, which is known to be more substantial than that of severe acute respiratory syndrome (SARS)–related coronavirus, which was the cause of the SARS outbreak that first appeared in Guangdong Province in southern China in 2002 [26]. Another work based on a stochastic mathematical prediction model of infection dynamics claimed that, by reducing worldwide travel by 90%, the epidemic's spread could be significantly reduced [27].

Other studies have sought to understand the propagation of misinformation related to COVID-19. One study used an epidemic model to represent the spread of misinformation about COVID-19 on various social media platforms such as Twitter, Instagram, YouTube, Reddit, and Gab; the study showed that users interact and consume information differently on each platform [28]. In this regard, media platforms such as Facebook, YouTube, and Twitter claim to attempt to redirect people to reliable sources of medical information and, to this end, have established direct lines of communication with the Centers for Disease Control and Prevention and the World Health Organization [29].

Among the regional research, one article argued that fake online news in Japan has led to xenophobia toward patients and Chinese visitors [30]. Another study surveyed 300,000 online panel members in South Korea in 2015, when the Middle East respiratory syndrome outbreak was prevalent in this country [31]. This work found that, if public health officials' information is untrustworthy, people rely more on online news outlets and communicate more via social media.

More recently, a report showed that the public could not easily receive the information on COVID-19 shared by public health officials due to prevalent misinformation on fake cures and conspiracy theories [32]. This study showed that infodemics' harm varied from country to country depending on public confidence in authorities. One study compared trends in three countries (ie, the United States, the United Kingdom, and Canada) in terms of political bias and found that, although political polarization surrounding COVID-19 exists in the United States and Canada, individuals' exact perspective on the pandemic is broadly related to the quality of their reasoning skills, regardless of political ideology [33].

Several studies have used data gathered from Twitter to analyze risk communication amid COVID-19. Some of them focus on sentiment analysis based on conventional rule-based lexicon models [34] or deep learning classifiers [35]. These studies measured the degree of sentiment polarity, such as positive and negative, and provided insights from observing daily sentiment changes.

Many types of data sets have been released to the public and research communities on COVID-19. One study crawled Twitter for approximately 3 months and collected information on tweets with relevant keywords in 10 languages [36]. Another work collated over 59,000 academic articles, including over 47,000 research papers, on COVID-19, SARS-CoV-2, and coronavirus-related issues [37] to conduct a comparison study.

Topic Modeling–Based Natural Language Processing

Natural language processing such as topic modeling is increasingly used to process extensive documents and extract

hidden thematic patterns of textual information [38]. Many studies have explored the capability of topic modeling in understanding the most important subjects of discussion on social media during crises and global epidemics such as dengue [39], Sika virus [40], and Ebola virus [41]. Given the remarkable performance of topic modeling in previous investigations, recent studies on the COVID-19 outbreak have also applied topic modeling to documents collected from different social media sites such as Facebook [42], Weibo [43], and Twitter [44].

One work analyzed COVID-19–related tweets over 2 weeks to study ongoing topics and found that Twitter can be considered a rich medium to understand public opinions in real time [45]. Another work conducted topic modeling on tweets to discover daily hot topics on the pandemic [46]. Furthermore, scholars leveraged Twitter to study the ecosystems of misinformation and conspiracy. One study has shown that users' political orientation correlates to their contribution to the spread of pandemic-related conspiracy [47]; another study demonstrates the link between fake news exposure and low trust in media [48]. Several techniques have been developed to detect conspiracy and misinformation on social media [49-51].

Despite the growing literature on risk communication during COVID-19, most studies that use topic modeling extract topics from either the entire studied period or manually segmented periods. This study considers time and topics jointly; we used an algorithmic approach to identify topical phases that arise naturally. Our goal is to observe changing risk communication contexts (even when conversations contain similar keywords) from the issue-attention cycle perspective. We also chose to study risk communication in Asian countries that have received relatively little attention. Our data method is not restricted to the studied countries; it can be applied to other languages and countries.

Methods

Data

We crawled Twitter for messages by using the Twint Python library [52] and search application programming interfaces [53]. Our analyses focused on four Asian countries (ie, South Korea, Iran, Vietnam, and India). We can ignore possible cultural differences in social media behaviors between Western and Asian users [54,55]. A common platform, Twitter, was used to study public conversations in these countries. Although multiple platforms exist, the open data access and global popularity make Twitter an appropriate medium to conduct a cross-national study.

The four countries were selected as a case study to demonstrate differences in their COVID-19 developments. In Iran, confirmed cases have gradually increased. In contrast, the case count in Vietnam has consistently stayed low. There was an abrupt increase in the numbers after the first confirmed case in South Korea, but the rising curve of confirmed cases has since flattened, unlike other countries. In India, the situation was relatively mild until mid-March 2020, and since then, there has been a drastic surge. Future research can replicate our methodology in other countries.

We set up two keywords, *corona* and *Wuhan pneumonia*, to crawl tweets and collected tweets for the 3 months from January to March 2020. In the studied countries, any tweets containing the official term *COVID-19* in local languages will be searched with the word *corona* (eg, in Korean, it is called *corona-19*). We also added *Wuhan* to collect unofficial terms of the virus. Table 1 lists the keywords used to collect data for each country. Keywords were decided after interviewing multiple local Twitter users for each country.

Table 1. Statistics of the scraped tweets.

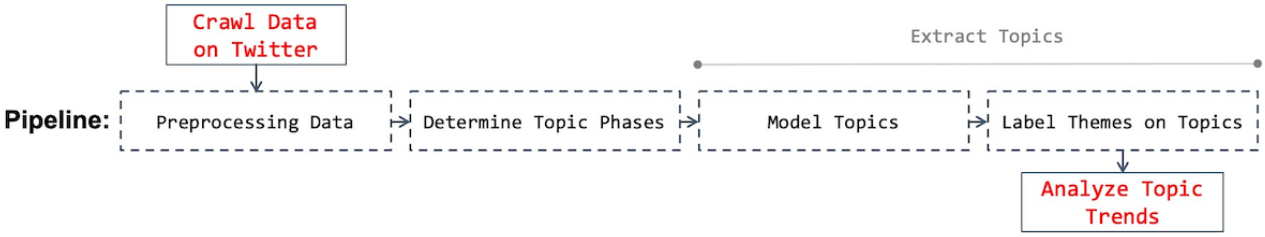
Language	Duration	Keywords[a] used	Tweets, n
Korean	January 1 to March 27, 2020	corona, Wuhan pneumonia	1,447,489
Farsi	January 1 to March 30, 2020	#corona, #coronavirus, #Wuhan, #pneumonia	459,610
Vietnamese	January 1 to March 31, 2020	corona, n-CoV, COVID, acute pneumonia	87,763
Hindi	January 1 to March 27, 2020	corona, Wuhan pneumonia	1,373,333

[a]Keywords were used to collect relevant data for each country. We used two kinds of keywords: one official naming of COVID-19 and *Wuhan* as an unofficial representative naming of the virus. Keywords listed here are translated in English from the actual local language (eg, "코로나" means "Corona" in Korean). The original keywords in local languages are listed in Table MA1 in Multimedia Appendix 1.

Demarcating Topic Phases and Extracting Topics

As shown in Figure 1, the data collection step was followed by the four modules for the extraction and labeling of major topics for certain phases. These steps were repeated for all four countries.

Figure 1. The pipeline structure of the topic analysis.



Step 1: Preprocessing Data

We first tokenized the data, a process that can be defined as converting data to the smallest units that have meaning. We filtered unnecessary textual information such as stop words, special characters (nonletters), special commands, and emojis. We then used existing Python tokenizer libraries corresponding to each language. Detailed information about the language-specific tokenizers is explained on GitHub [16].

Step 2: Determine Topic Phases

The next step is to demarcate specific phases divided by dates to extract topics. This is nontrivial since there are multiple fluctuations and changes in topics reflecting real events such as increased patients with COVID-19. Furthermore, we ruled out using the epidemic phases announced by each government because the offline epidemic phases do not seem to capture actual online topic trends as explained in the forthcoming Basic Daily Trends section.

The issue-attention cycle moderating public attention to a given issue can be measured in media attention, such as the number of news stories [17-19,21-23] or tweets [20,25] on the topic. We, therefore, isolated dates that show sudden increases in the daily tweet volume. We set up two learnable parameters of the first derivatives (hereafter *velocity*) and the second derivatives (hereafter *acceleration*) of the daily tweet volumes, as illustrated in the following equations, where D is a day, t is a target date, and $t-1$ is 1 day past t:

We set the *velocity* and *acceleration* values when *the country announced the first confirmed case* as the ground truth (GT). This approach's intuition is that *velocity* and *acceleration* are proxies for each country's unique communication traits regarding a specific subject (ie, COVID-19 in our case). Once these values were computed from the first confirmed date, they were set identically for the remaining periods.

We established joint thresholds for *velocity* and *acceleration* to find dates where *velocity* is still smaller than $velocity_{GT}$, and *acceleration* becomes more substantial than $acceleration_{GT}$ ($0<velocity<velocity_{GT}$ and $acceleration>acceleration_{GT}$). In this manner, we identified the two parameters from the date of the first confirmed case by country and then detected other dates conjectured to be the start of forthcoming topic phases. When learning these parameters, for *velocity*, we rounded down the $velocity_{GT}$ value and added 1, and for *acceleration*, we rounded down *accelerationGT*, which is similar to the machine learning approach's concept of loss minimization (ie, a learning process is finished by one step).

We adopted a low-pass filter with 0.2 as the low-frequency threshold to remove noisy signals and smooth the data. Finally, the temporal data are divided into topic phases (see Multimedia Appendix 1 to find the computed daily *velocity* and *acceleration* trends and demarcated phases by country).

Step 3: Extract Topics—Model Topics

We used latent Dirichlet allocation (LDA) for the topic modeling task. LDA is a well-known machine learning method to extract topics from given textual documents (ie, a collection of discrete data points) [56]. LDA generates and maximizes the joint probability of the topics' word distribution and the documents' topic distribution [38]. For short sentences, LDA-based methods may not work correctly due to sparse co-occurrences. However, in our case, tweets were collected via specific keywords; therefore, the topics were the focus, and the word co-occurrences among tweets were no longer sparse. Various studies have used the same LDA method on short documents, including Twitter [57-59].

The topic count for each phase is a hyperparameter. The topic count's range is between 2 and 50. We calculated perplexity, that is, the probability of how many tokens might be placed in the next step (ie, indicating ambiguity over the next possible token). Perplexity is a metric that is often used to optimize language models [60]. The minimum required frequency of words for each phase in tweets was set to 20, and each phase's epoch (ie, a number of iterations to train LDA) was set to 100. We then decided the optimum number of topics for each phase by choosing the minimum perplexity value. We further analyzed our modeling results' reliability and confirmed that the results were steady and dependable (see Multimedia Appendix 2 for more details).

Table 2 shows how many prominent topical phases were found for each country. For each phase, we list the statistics of the risk communication, including the period of the topical phase, the total tweet count during the phase, the average user count per day, the average original and retweet counts per day, the ratio of original tweets to retweets, and the number of topics suggested by perplexity.

Table 2. The optimal number of phases and topics by country.

Country	Phase 0	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
South Korea					N/A[a]	N/A
Time period	Jan 1-19, 2020	Jan 20-Feb 12, 2020	Feb 13-Mar 9, 2020	Mar 10-27, 2020		
Total tweets, n	507	161,790	672,080	366,073		
Average users per day	14.06	2415.52	5376.77	5577.88		
Average original tweets per day	28.17	5244.09	17,796.08	13,095.65		
Average retweets per day	21.78	56,809.78	211,310.89	147,759.41		
Tweet depth[b]	0.77	10.83	11.87	11.28		
Topics determined by perplexity, n	2	41	15	43		
75th percentile of topics[c], n	1	18	6	14		
Final topics[d], n	1	8	5	11		
Iran			N/A	N/A	N/A	N/A
Time period	Jan 1-Feb 18, 2020	Feb 19-Mar 30, 2020				
Total tweets, n	15,473	437,176				
Average users per day	245.34	1442.46				
Average original tweets per day	385.63	5272.04				
Average retweets per day	1315.13	22,128.76				
Tweet depth	3.41	4.20				
Topics determined by perplexity, n	3	5				
75th percentile of topics, n	2	4				
Final topics, n	3	6				
Vietnam						
Time period	Jan 1-20, 2020	Jan 21-25, 2020	Jan 26-Feb 15, 2020	Feb 16-Mar 4, 2020	Mar 5-22, 2020	Mar 23-31, 2020
Total tweets, n	140	1499	18,424	28,458	26,950	12,292
Average users per day	3.79	131.25	179.65	485.59	340.65	433.29
Average original tweets per day	7.37	218.50	686.60	1238.77	1089.94	1224.00
Average retweets per day	0.21	20.75	159.80	582.29	192.24	201.86
Tweet depth	0.03	0.09	0.23	0.47	0.18	0.16
Topics determined by perplexity, n	19	3	6	46	48	16
75th percentile of topics, n	1	1	3	22	19	4
Final topics, n	1	2	4	7	10	2
India				N/A	N/A	N/A
Time period	Jan 1-29, 2020	Jan 30-Mar 9, 2020	Mar 10-27, 2020			
Total tweets, n	3088	151,210	1,219,030			
Average users per day	107.41	1364.95	13,318.63			
Average original tweets per day	269.72	4261.13	58,924.55			
Average retweets per day	415.69	14,467.8	318,368.05			
Tweet depth	1.54	3.40	5.40			
Topics determined by perplexity, n	3	50	47			
75th percentile of topics, n	2	22	20			

Country	Phase 0	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
Final topics, n	3	5	9			

[a]N/A: not applicable.

[b]Measured as the ratio of retweets to original tweets.

[c]Major topics.

[d]After human annotators merged similar themes.

Step 4: Extract Topics—Label Topics

This step involves labeling the themes of the extracted topics and allocating semantic meanings to each topic. We first sorted all tweets with the identified topics in descending order (ie, tweets on the most prevalent topics listed first) and discarded the minor topics that accounted for less than 25% of all tweets.

We then extracted the top 1000 retweeted tweets and the 30 keywords with the highest probability of usage for each topic. We provided these data sets to local users from each country and asked them to label themes for each topic based on the given data sets. Any similar or hierarchical topics were then merged via qualitative coding into a higher category. If one topic corresponded to several themes, then it was given multiple class labels. The maximum number of multiple classes within topics was two, and each class within a topic was weighted as 0.5 in the plot of daily trends in the number of tweets.

Human annotators, who are familiar with the local language and Twitter, qualitatively assessed the extracted topics. First was the intralevel, where annotators labeled each topic based on the contents of the sampled top 1000 tweets and top 30 words. The second was the interlevel, where the annotators compared tweet contents and top-occurring words among topics regardless of the phase. Other annotators then cross-checked the assessment.

The Cohen kappa coefficient to measure the intercoder reliability was 0.766 (see Multimedia Appendix 2 for details on this validation and the list of topics and top-occurring words for each country). Our analysis objective was not to substitute human laborers on monitoring misinformation but to assist them by grouping tweets into specific topics, including misinformation.

Concerning the local and global news themes, we narrowed down the labels since people talked about different news categories. We sublabeled tweets as "_confirmed" if it was about confirmed cases or deaths, "_hate" if it was about hate crimes toward individual races, "_economy" if it was about the economic situation and economic policies, "_cheerup" if it was about supporting each other, and "_education" if it was about when to reopen schools; finally, no sublabel was given to tweets about general information.

Results

Basic Daily Trends

Figure 2 shows trends of the daily tweet count; the same trend is shown along with the number of confirmed COVID-19 cases for each country in Figures 3-6. Adding to the two trends, we included each government's official epidemic phases as vertical lines. It is evident in the figures that the tweet trends are associated with the confirmed cases. However, the official epidemic phases do not accurately explain the tweet trends. We examine trends for each country in the following sections.

Figure 2. Daily trends in the four countries. The x-axis is dated, and the y-axis is the number of tweets with a log scale.

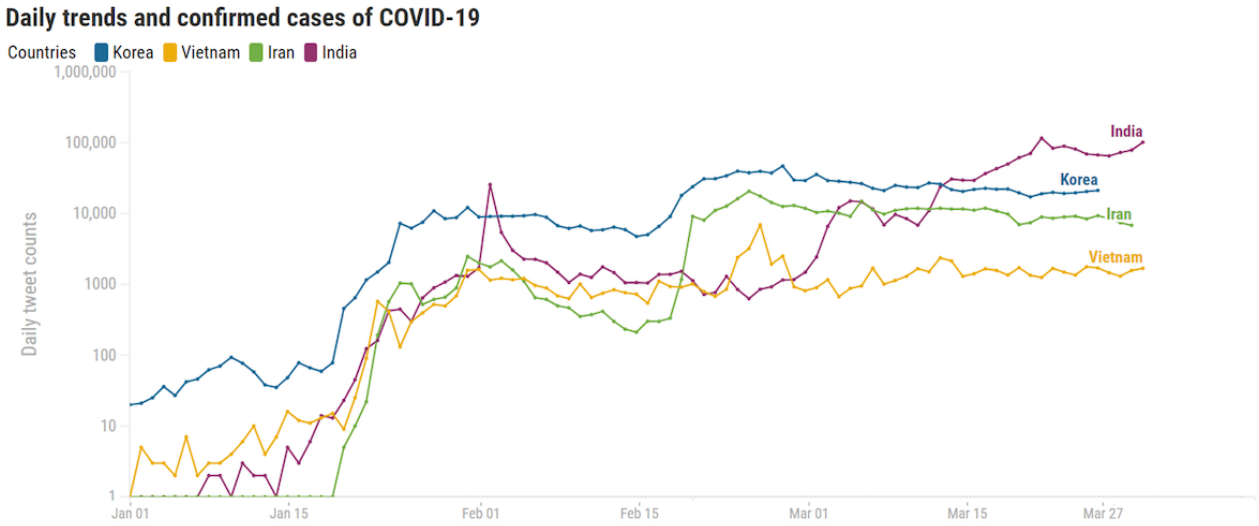


Figure 3. Daily trends in South Korea. Start/end dates of the official epidemic phases (vertical dashed lines), trends in the number of tweets (blue lines), and trends in the number of confirmed cases (red bars).

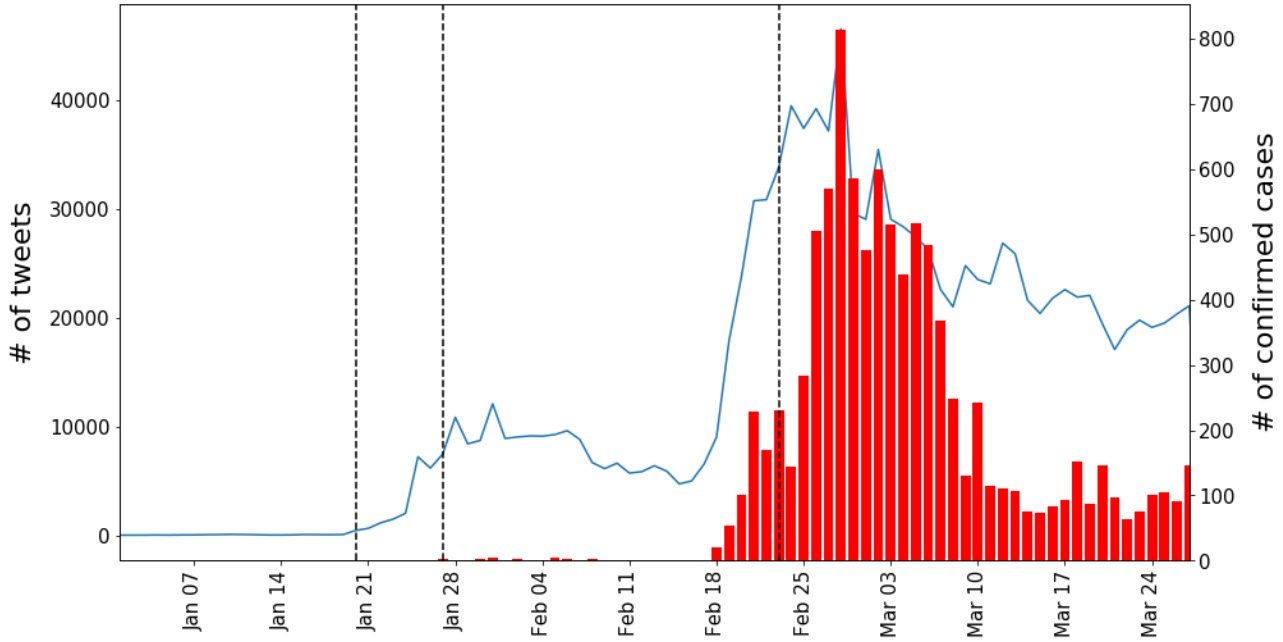


Figure 4. Daily trends in Iran. Start/end dates of the official epidemic phases (vertical dashed lines), trends in the number of tweets (blue lines), and trends in the number of confirmed cases (red bars).

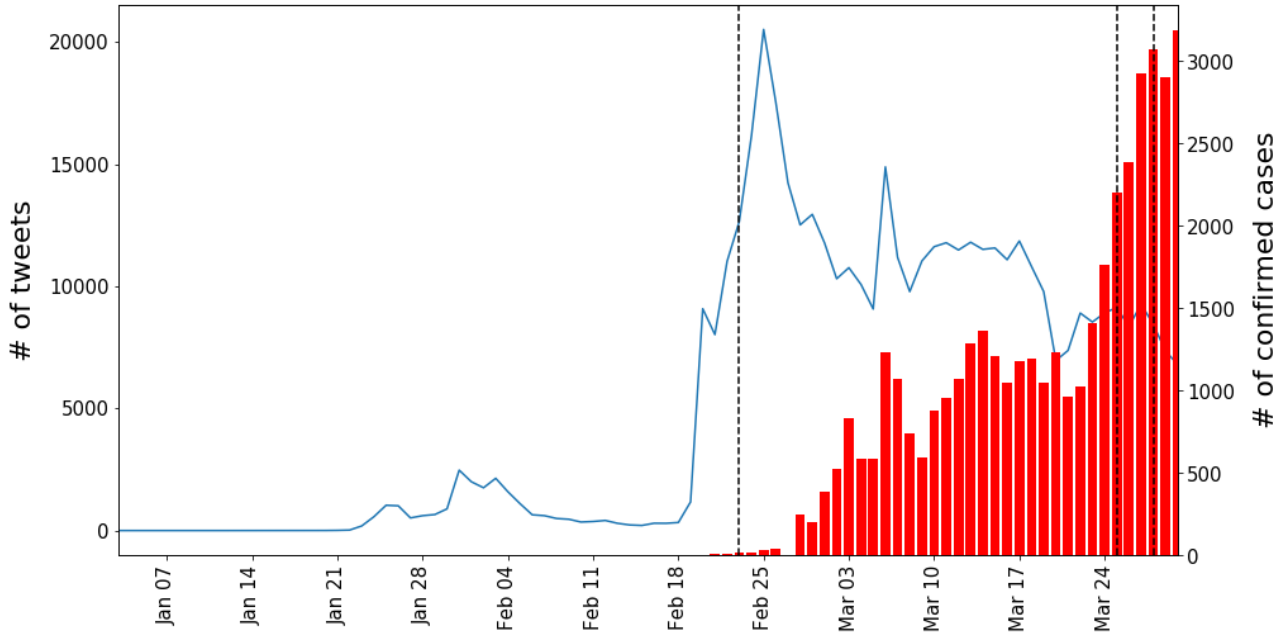


Figure 5. Daily trends in Vietnam. Start/end dates of the official epidemic phases (vertical dashed lines), trends in the number of tweets (blue lines), and trends in the number of confirmed cases (red bars).

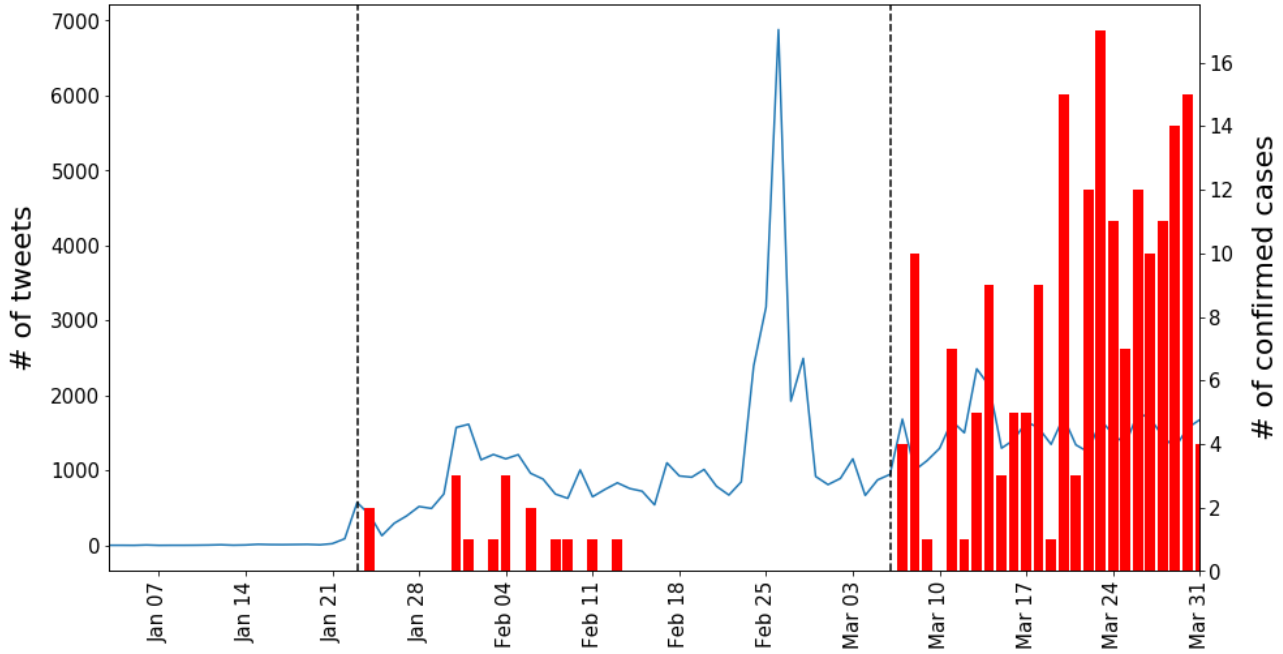
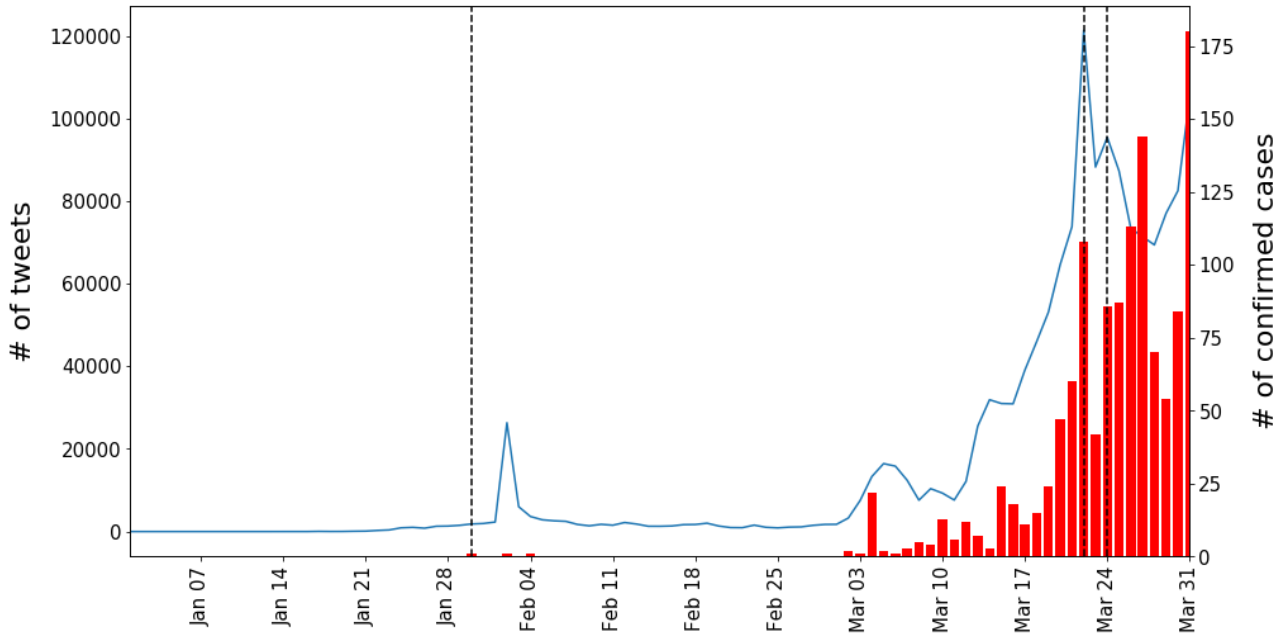


Figure 6. Daily trends in India. Start/end dates of the official epidemic phases (vertical dashed lines), trends in the number of tweets (blue lines), and trends in the number of confirmed cases (red bars).



South Korea

The first patient with COVID-19 was reported in South Korea on January 20, 2020. This explains why the tweet count remains relatively low during early January and mostly increases only after late January (see Figure 3). On January 25, the Korean government issued a travel warning for Wuhan and Hubei Province, and suggested that Korean citizens evacuate from those areas, which was heavily discussed on Twitter.

On February 18, 2020, the tweet numbers increased sharply due to the 31st confirmed case related to a cult religious group Shincheonji in Daegu City. After this case was confirmed, the quarantine authority began rigorous testing, focusing on Daegu, and the number of confirmed cases increased drastically until mid-March. The tweet trends follow an identical pattern. However, the official epidemic phases announced by the government, represented by vertical dashed lines in the figure, seem to lag behind the increases in the number of tweets. This pattern shows that the official epidemic phases do not align well with the amount of online attention.

Other Countries

We repeated the analysis with the other three countries, as shown in Figures 4-6 (see Multimedia Appendix 3 for each country's detailed explanation).

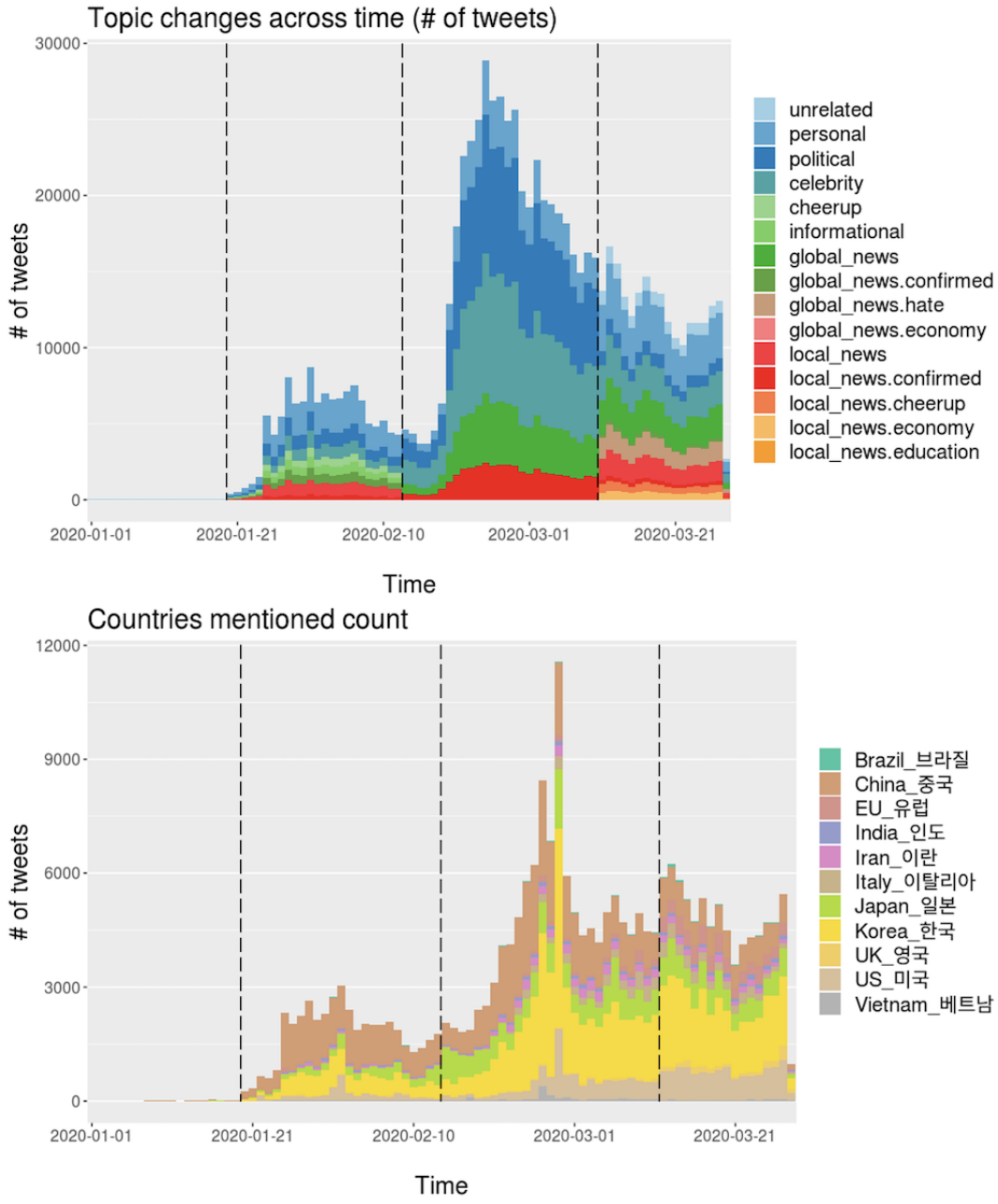
Extracted Topic Trends

We used the daily theme labels acquired from the "Label Topics" module and analyzed the topic changes over time with plots for the four countries. One plot showed daily trends based on the number of tweets, while another plot shows trends based on the number of tweets mentioning country names such as the United States. Overall, as people talked more about the COVID-19 outbreak (ie, as the daily number of tweets increased), people's topics became less diverse.

South Korea

The data yielded a total of four topic phases, which are used in Figure 7. Phase 0 has no related topics. For phases 1, 2, and 3, the number of topics varies from 8, 5, and 11, respectively. In phase 1, people talked a great deal about their personal thoughts and opinions linked to the current outbreak, and they tried to cheer each other up. In phase 2, as the crisis peaked, people talked less about personal issues and mainly about political and celebrity issues. In Korea, political discussions revolved around closing the South Korean border with China and other countries. In phase 3, as the daily number of tweets decreased relative to that in phase 2, people talked about diverse topics, including local and global news. The major topics here included worries about hate crimes directed toward Asians in Western countries. Such diverse topics are likely shown when people think the pandemic has just passed its peak.

Figure 7. Daily topic trends in South Korea. Trends based on number of tweets (top) and based on number of tweets mentioning country names (bottom).



We portrayed daily trends of interest in other countries by counting the tweets mentioning other countries' names in local languages or English. Korea, China, and Japan were mentioned most frequently; we suspect that this was mainly triggered by political and diplomatic relationships. Meanwhile, the United States and Italy were both mentioned steadily across the 3 months, with the media outlets broadcasting global news affecting this phenomenon.

Other Countries

We repeated the same analysis and interpreted the results for the other cases (Iran, Vietnam, and India), as depicted in Figures 8-10 (see Multimedia Appendix 2 for the derived topic trend and the mentioned country name trend graphs and detailed corresponding explanations by country).

Figure 8. Daily topic trends in Iran based on the number of tweets.

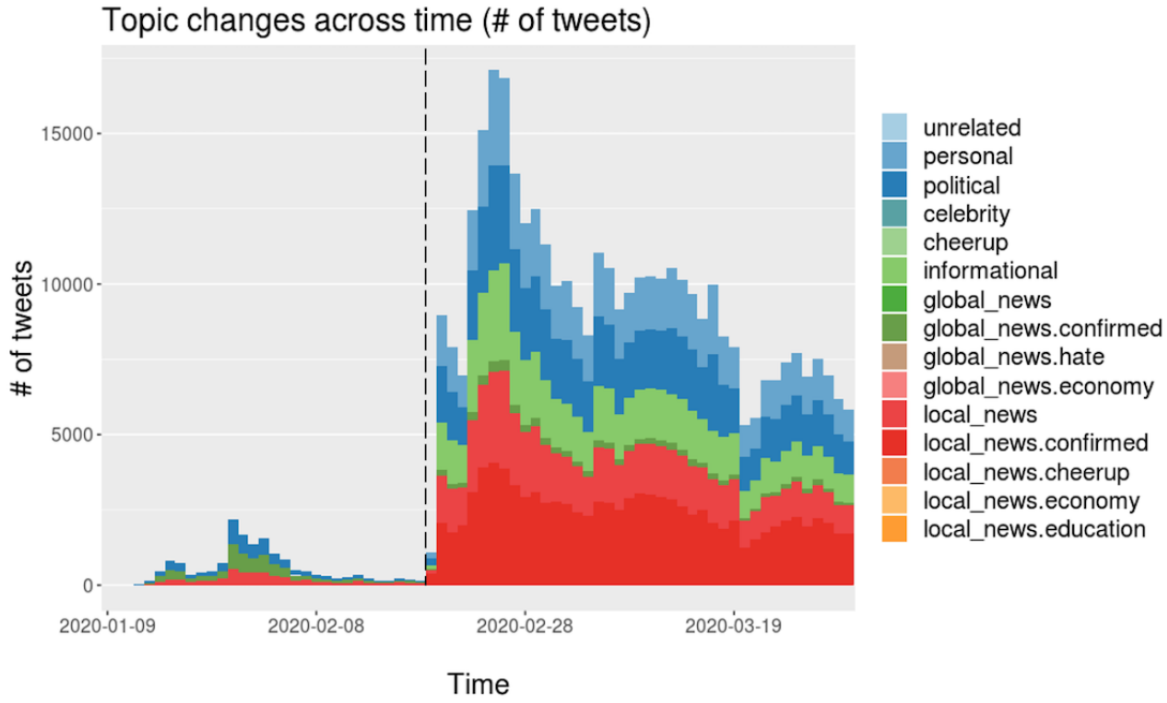


Figure 9. Daily topic trends in Vietnam based on the number of tweets.

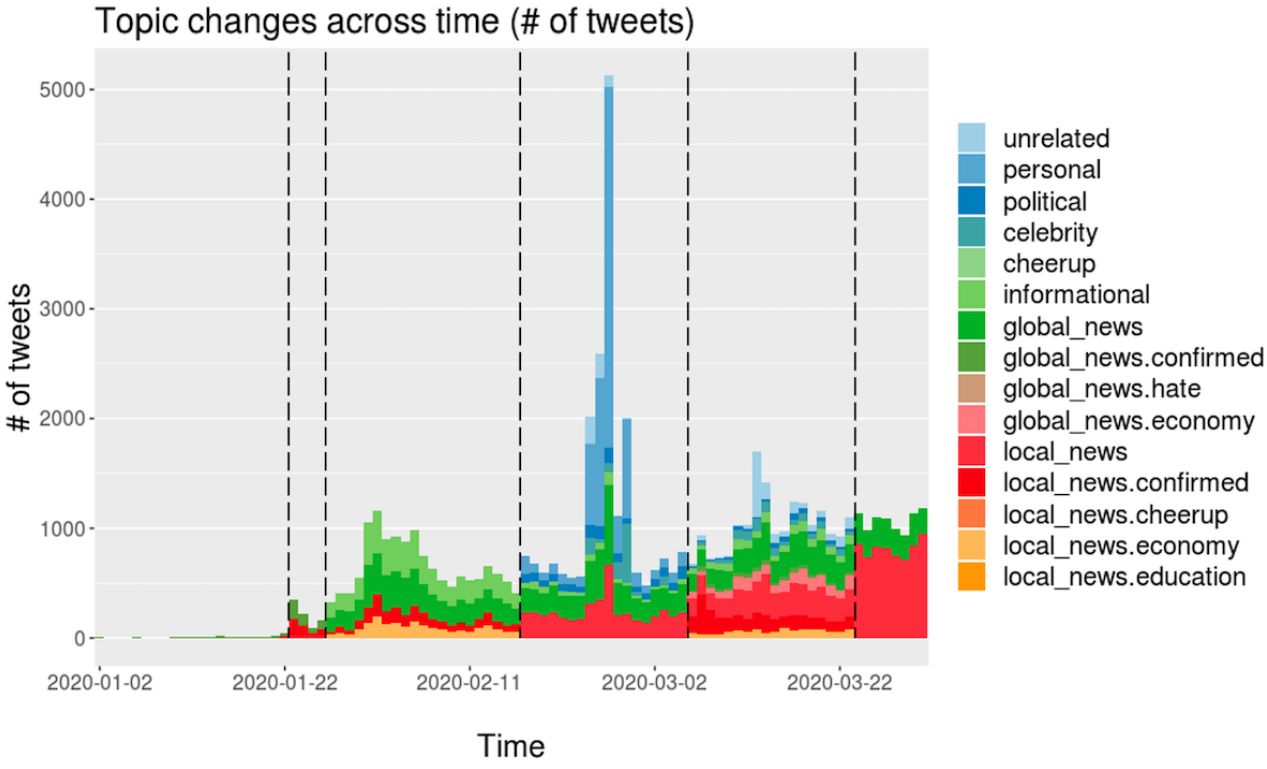
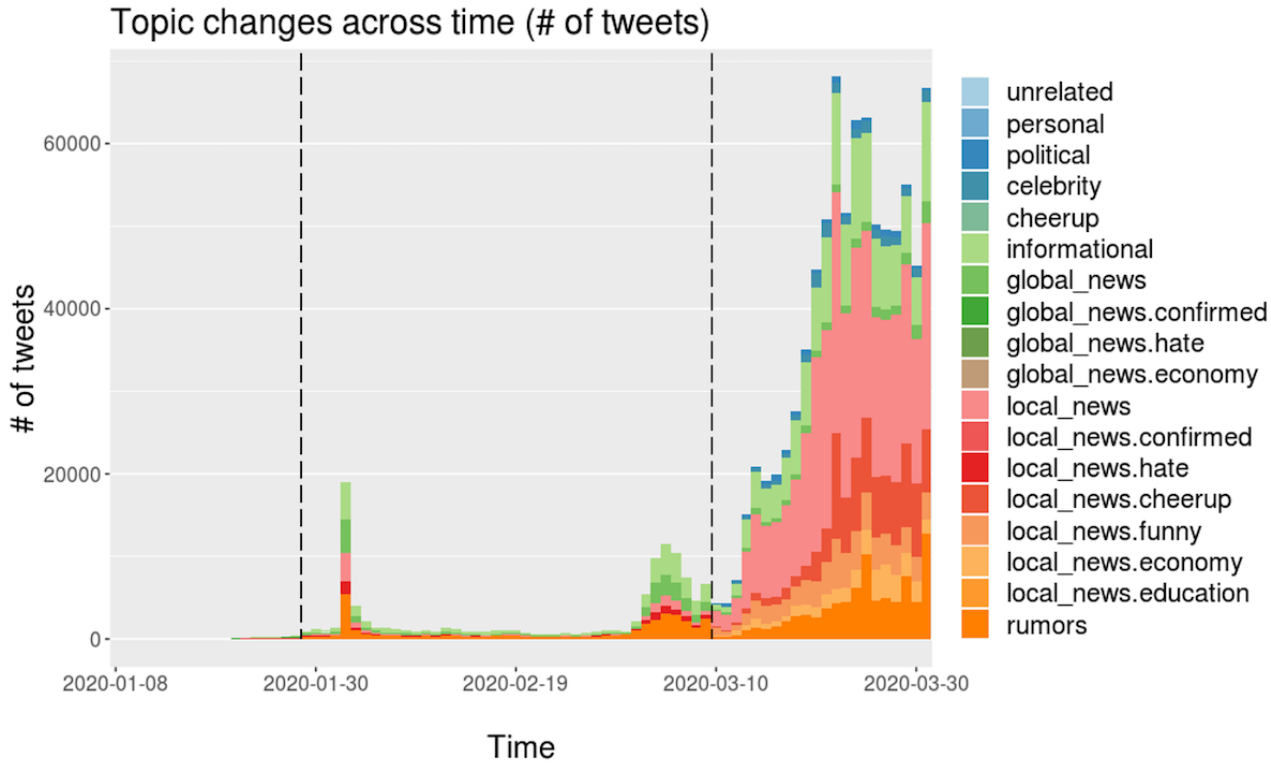


Figure 10. Daily topic trends in India based on the number of tweets.



Discussion

RQ 1 and RQ 2: Explore an Automatic Way to Decide Topic Phase and Model Topics

This paper analyzes tweets to understand the public discourse on the COVID-19 pandemic. In South Korea, the daily numbers of tweets reached their local maxima in tandem with major offline events. However, in Iran and Vietnam, the tweet counts did not synchronize well with offline events; this may be because of various reasons (eg, Twitter is only one of the platforms used by citizens of this country). Overall, it is interesting to observe that the Twitter data peaks do not necessarily correlate with local governments' announcements. Social media attention can precede the official announcements, while the official announcements can reinforce the attention.

RQ 3: Explore Common Traits Among Countries on Risk Communication

Based on the topics labeled as people talked more about COVID-19, they tended to refer to a smaller number of topics. This was more apparent when the tweet depth value was used for the phases, as presented in Table 2.

Tweet depth is defined as the number of retweets per day divided by the number of tweets per day. It can be deemed a measure of standardized cascading depth, with a higher value signifying a greater depth for one tweet. The country-level sociopolitical and cultural background, and Twitter popularity may lead to the observed differences in tweet depth. We verified that tweet depth tended to increase in South Korea and Vietnam cases when people communicated more about COVID-19. This phenomenon reaffirms the finding in another study that the online coronavirus network's diameter value was smaller than that of other keyword networks [61].

The topical phases with the most considerable tweet depth appeared in the second stage of the issue-attention cycle, where public awareness of an issue soars. In Iran and India, the number of phases might have been too small to discern any such trends. It is also worth noting that this pattern has no intercountry temporal dependence. In other words, even though the pandemic hit the countries at different times, our analysis shows that the tweet depth reached a maximum when the pandemic worsened in that country. This observation could prove to be an effective forewarning of upcoming misinformation cascades.

Moreover, the daily tweet volume peaks reflected the daily number of confirmed cases. In Iran, Vietnam, and India, the daily tweet volume peak anticipated the peak of the number of daily confirmed cases by up to a few weeks. Although the two peaks are close to each other for South Korea, it is worth noting that, around the time of their occurrence, South Korea was becoming the country most affected by COVID-19 outside mainland China.

Interestingly, as shown in Figures 3-6, a simultaneous upsurge in the numbers of tweets occurred in South Korea, Iran, and Vietnam (but not in India) at the end of February 2020, before the upsurge in numbers of locally confirmed cases. Given that COVID-19 is a global issue, this suggests that the issue-attention cycle on a social media platform is more responsive to global rather than local events. In this light, the COVID-19 pandemic offers a gripping opportunity for future researchers to theorize the issue-attention cycle model on a global scale and see how the cycle evolves in conjunction with location-specific topics such as increasing or decreasing numbers of confirmed cases, government measures, and social conflicts.

RQ 4: Explore Unique Traits by Countries on Risk Communication

We also observed a number of countrywise differences. One of them is the national versus international focus of South Korea and Vietnam during the initial phase. Phase 0 tweets in Korea were not directly related to COVID-19 but simply contained the word *corona* in a different context. This is because this time period was before the first public announcement of the confirmed patients in Korea. In contrast, in Vietnam, the first phase tweets were concerned with international updates on COVID-19. The difference is likely explained by the increasing patient count worldwide. Note that South Korea was one of the first countries to experience the pandemic. We did not attempt to draw any general conclusions from these findings due to the small tweet volumes in phase 0 for both countries. Nevertheless, Vietnamese users discussed the global epidemic more than Korean users from the outset. This tendency may have been associated with Vietnam's successful defense against the pandemic later on.

With specific reference to each country, in South Korea, when the local (offline) pandemic situation became severe (phase 2), the number of topics discussed on Twitter decreased, which means that people focused more on only a handful of issues. A unique feature of phase 0 was that people sought to cheer each other up and express solidarity in difficult times. In Iran's case, the topic count was relatively steady over time. The significant topics discussed were confined to news and information; we interpreted this as a sign that Iranian users tend to be cautious about using social media.

For Vietnam, in phase 4, when tweet traffic was lower than in phase 3, the number of topics became more substantial, and the topic themes became less related to the numbers of confirmed cases and death tolls. For instance, people talked more about the economy in phases 2 and 4. The Indian case also displayed a unique trait: many topics were related to misinformation, the scale of which was much lower in the other countries. A large portion of the topics consisted of misinformation and hateful content; this trend was observed throughout phases 2 and 3 (see Multimedia Appendix 2).

Limitations and Future Work

There are several limitations to be considered. First, we analyzed tweets from only four countries, and therefore, we need to be cautious about extrapolating explanations and insights generally. We plan to extend this study by including more countries. Second, there are other ways to demarcate the topic phases. Our approach was informed by the issue-attention cycle framework, as we computed unique communication traits (ie, *velocity* and *acceleration* by country) that should be relatively consistent across nations throughout the COVID-19 pandemic.

Last, there are also other methodologies to model topics. One natural extension would be to use the external web links that are embedded in the relevant tweets. Scraping the content from external web pages could provide richer contexts in understanding risk communication on social media. One recent work used multilingual Bidirectional Encoder Representations from Transformers, a well-known transformer-based deep embedding model, and fine-tuned it by considering topical and temporal information to model topics of COVID-19 tweets [62]. On deciding topic phases via the data itself, one may use LDA and other embedding methods to model topics.

Concluding Remarks

The current literature on the infodemic has emphasized the social media platform's content moderation efforts [63] and fact-checking as a key risk communication strategy [64,65]. This study extends these scholarly endeavors. Predicated on an issue-attention cycle framework, we analyzed public attention on COVID-19–related topics in four Asian countries. We used a time-topic cohesive approach to automatically identify transitions in topical interests and qualitatively evaluated the topics found by local users.

Our research found that when the tweet count on COVID-19 increased, it did not lead to an increased number of topics; regardless of the tweet count, much of the public attention remained focused on a limited set of topics. The early days of the COVID-19 pandemic also involved various misinformation and hateful speech in the studied countries; fake news was one of the central topics discussed (not a peripheral topic). The proposed steps could indicate the global effects of infodemics during a pandemic and identify the emergence of misinformation and its prevalence, which will help prioritize which misinformation to debunk.

Acknowledgments

The authors would like to thank Tae-Gwan Kang for his insightful comments. SP, SH, and MC were supported by the Institute for Basic Science (IBS-R029-C2) and the Basic Science Research Program through the National Research Foundation of Korea (No. NRF-2017R1E1A1A01076400).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data/code description, computed daily velocity/acceleration trends for each country, and derived temporal phases.
[PDF File (Adobe PDF File), 494 KB - jmir_v23i3e23272_app1.pdf]

Multimedia Appendix 2
Daily topic trends on social media, the reliability of the topic modeling results, intercoder reliability, and the labeled list of major topics by country.
[PDF File (Adobe PDF File), 1904 KB - jmir_v23i3e23272_app2.pdf]

Multimedia Appendix 3
Daily numbers of COVID-19 confirmed cases and tweet trends by country.
[PDF File (Adobe PDF File), 347 KB - jmir_v23i3e23272_app3.pdf]

References

1. Ahmad AR, Murad HR. The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: online questionnaire study. J Med Internet Res 2020 May 19;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]
2. Beaunoyer E, Dupéré S, Guitton M. COVID-19 and digital inequalities: reciprocal impacts and mitigation strategies. Comput Human Behav 2020 Oct;111:106424 [FREE Full text] [doi: 10.1016/j.chb.2020.106424] [Medline: 32398890]
3. Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. Science 2018 Mar 09;359(6380):1094-1096. [doi: 10.1126/science.aao2998] [Medline: 29590025]
4. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
5. Kwon S, Cha M, Jung K, Chen W, Wang Y. Prominent features of rumor propagation in online social media. 2013 Presented at: 2013 IEEE 13th International Conference on Data Mining; December 7-10, 2013; Dallas, TX. [doi: 10.1109/icdm.2013.61]
6. Coronavirus disease 2019 (COVID-19): situation report – 86. World Health Organization. 2020. URL: https://bit.ly/2SKCl8X [accessed 2020-08-27]
7. Stecula DA, Kuru O, Hall Jamieson K. How trust in experts and media use affect acceptance of common anti-vaccination claims. Harvard Kennedy Sch Misinformation Rev 2020 Jan 14;1(1):1-11. [doi: 10.37016/mr-2020-007]
8. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
9. Renn O, Leiss W, Kastenholz H. OECD Guidance Document on Risk Communication for Chemical Product Risks. Paris: OECD; 2002.
10. Cha M, Cha C, Singh K, Lima G, Ahn Y, Kulshrestha J, et al. Prevalence of misinformation and factchecks on the COVID-19 pandemic in 35 countries: observational infodemiology study. JMIR Hum Factors 2021 Feb 13;8(1):e23279 [FREE Full text] [doi: 10.2196/23279] [Medline: 33395395]
11. Cheng T, Wicks T. Event detection using Twitter: a spatio-temporal approach. PLoS One 2014;9(6):e97807 [FREE Full text] [doi: 10.1371/journal.pone.0097807] [Medline: 24893168]
12. Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, et al. A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. J Med Internet Res 2014 Oct 20;16(10):e236 [FREE Full text] [doi: 10.2196/jmir.3416] [Medline: 25331122]
13. Farnaghi M, Ghaemi Z, Mansourian A. Dynamic spatio-temporal tweet mining for event detection: a case study of Hurricane Florence. Int J Disaster Risk Sci 2020 May 26;11(3):378-393. [doi: 10.1007/s13753-020-00280-z]
14. Malizia N. Inaccuracy, uncertainty and the space-time permutation scan statistic. PLoS One 2013;8(2):e52034 [FREE Full text] [doi: 10.1371/journal.pone.0052034] [Medline: 23408930]
15. Downs A. Up and down with ecology—the "issue-attention cycle". Public Interest 1972;28:38-51.
16. dscig / COVID19_tweetsTopic. GitHub. 2020. URL: https://github.com/dscig/COVID19_tweetsTopic [accessed 2021-03-12]
17. Nisbet MC, Huge M. Attention cycles and frames in the plant biotechnology debate: managing power and participation through the press/policy connection. Int J Press/Politics 2016 Sep 14;11(2):3-40. [doi: 10.1177/1081180x06286701]
18. McComas K, Shanahan J. Telling stories about global climate change: measuring the impact of narratives on issue cycles. Commun Res 2016 Jun 30;26(1):30-57. [doi: 10.1177/009365099026001003]
19. Anderson AA, Brossard D, Scheufele DA. News coverage of controversial emerging technologies. Evidence for the issue attention cycle in print and online media. Politics Life Sci 2012;31(1-2):87-96. [doi: 10.2990/31_1-2_87] [Medline: 23379318]
20. Wang W, Guo L. Framing genetically modified mosquitoes in the online news and Twitter: intermedia frame setting in the issue-attention cycle. Public Underst Sci 2018 Nov;27(8):937-951. [doi: 10.1177/0963662518799564] [Medline: 30198824]
21. Shih T, Wijaya R, Brossard D. Media coverage of public health epidemics: linking framing and issue attention cycle toward an integrated theory of print news coverage of epidemics. Mass Commun Soc 2008 Apr 07;11(2):141-160. [doi: 10.1080/15205430701668121]
22. Arendt F, Scherr S. Investigating an issue-attention-action cycle: a case study on the chronology of media attention, public attention, and actual vaccination behavior during the 2019 measles outbreak in Austria. J Health Commun 2019;24(7-8):654-662. [doi: 10.1080/10810730.2019.1652709] [Medline: 31423919]

23. Jung Oh H, Hove T, Paek H, Lee B, Lee H, Kyu Song S. Attention cycles and the H1N1 pandemic: a cross-national study of US and Korean newspaper coverage. Asian J Commun 2012 Apr;22(2):214-232. [doi: 10.1080/01292986.2011.642395]
24. David CC, Ong JC, Legara EFT. Tweeting Supertyphoon Haiyan: evolving functions of Twitter during and after a disaster event. PLoS One 2016;11(3):e0150190 [FREE Full text] [doi: 10.1371/journal.pone.0150190] [Medline: 27019425]
25. Jang SM, Park YJ, Lee H. Round-trip agenda setting: tracking the intermedia process over time in the ice bucket challenge. Journalism 2016 Aug 30;18(10):1292-1308. [doi: 10.1177/1464884916665405]
26. Liu Y, Gayle A, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med 2020 Mar 13;27(2):taaa021 [FREE Full text] [doi: 10.1093/jtm/taaa021] [Medline: 32052846]
27. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
28. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 06;10(1):16598. [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]
29. Frenkel S, Alba D, Zhong R. Surge of virus misinformation stumps Facebook and Twitter. The New York Times. 2020 Mar 08. URL: https://nyti.ms/2AI5Ylw [accessed 2020-08-27]
30. Shimizu K. 2019-nCoV, fake news, and racism. Lancet 2020 Feb 29;395(10225):685-686 [FREE Full text] [doi: 10.1016/S0140-6736(20)30357-3] [Medline: 32059801]
31. Jang K, Baek YM. When information from public health officials is untrustworthy: the use of online news, interpersonal networks, and social media during the MERS outbreak in South Korea. Health Commun 2019 Aug;34(9):991-998. [doi: 10.1080/10410236.2018.1449552] [Medline: 29558170]
32. Oxford Analytica. Misinformation will undermine coronavirus responses. Expert Briefings 2020:1. [doi: 10.1108/oxan-db250989]
33. Pennycook G, McPhetres J, Bago B, Rand DG. Predictors of attitudes and misperceptions about COVID-19 in Canada, the UK, and the USA. psyArXiv. Preprint posted online April 14, 2020 [FREE Full text]
34. Zhou J, Yang S, Xiao C, Chen F. Examination of community sentiment dynamics due to Covid-19 pandemic: a case study from Australia. arXiv. Preprint posted online June 22, 2020 [FREE Full text]
35. Imran AS, Daudpota SM, Kastrati Z, Batra R. Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets. IEEE Access 2020;8:181074-181090. [doi: 10.1109/access.2020.3027350]
36. Chen E, Lerman K, Ferrara E. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. JMIR Public Health Surveill 2020;6(2):e19273 [FREE Full text] [doi: 10.2196/19273] [Medline: 32427106]
37. Wang LL, Lo K, Chandrasekhar Y, Reas R, Yang J, Eide D, et al. CORD-19: the Covid-19 Open Research Dataset. arXiv. Preprint posted online April 22, 2020 [FREE Full text] [Medline: 32510522]
38. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Machine Learning Res 2003;3:993-1022.
39. Missier P, Romanovsky A, Miu T, Pal A, Daniilakis M, Garcia A, et al. Tracking dengue epidemics using Twitter content classification and topic modelling. In: Casteleyn S, Dolog P, Pautasso C, editors. Current Trends in Web Engineering ICWE 2016 International Workshops, DUI, TELERISE, SoWeMine, and Liquid Web, Lugano, Switzerland, June 6-9, 2016. Revised Selected Papers. Cham: Springer; 2016:80-92.
40. Miller M, Banerjee T, Muppalla R, Romine W, Sheth A. What are people tweeting about zika? An exploratory study concerning its symptoms, treatment, transmission, and prevention. JMIR Public Health Surveill 2017 Jun 19;3(2):e38 [FREE Full text] [doi: 10.2196/publichealth.7157] [Medline: 28630032]
41. Lazard AJ, Scheinfeld E, Bernhardt JM, Wilcox GB, Suran M. Detecting themes of public concern: a text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. Am J Infect Control 2015 Oct 01;43(10):1109-1111. [doi: 10.1016/j.ajic.2015.05.025] [Medline: 26138998]
42. Sesagiri Raamkumar A, Tan SG, Wee HL. Measuring the outreach efforts of public health authorities and the public response on Facebook during the COVID-19 pandemic in early 2020: cross-country comparison. J Med Internet Res 2020 May 19;22(5):e19334 [FREE Full text] [doi: 10.2196/19334] [Medline: 32401219]
43. Han X, Wang J, Zhang M, Wang X. Using social media to mine and analyze public opinion related to COVID-19 in China. Int J Environ Res Public Health 2020 Apr 17;17(8):2788 [FREE Full text] [doi: 10.3390/ijerph17082788] [Medline: 32316647]
44. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
45. Medford RJ, Saleh SN, Sumarsono A, Perl TM, Lehmann CU. An “Infodemic”: leveraging high-volume Twitter data to understand public sentiment for the COVID-19 outbreak. medRxiv. Preprint posted online April 7, 2020. [doi: 10.1101/2020.04.03.20052936]
46. Yin H, Yang S, Li J. Detecting topic and sentiment dynamics due to Covid-19 pandemic using social media. In: Yang X, Wang CD, Islam MS, Zhang Z, editors. Advanced Data Mining and Applications. ADMA 2020. Lecture Notes in Computer Science, vol 12447. Cham: Springer; 2021:610-623. [doi: 10.1007/978-3-030-65390-3_46]

47. Shahrezaye M, Meckel M, Steinacker L, Suter V. COVID-19's (mis) information ecosystem on Twitter: how partisanship boosts the spread of conspiracy narratives on German speaking Twitter. arXiv. Preprint posted online September 27, 2020 [FREE Full text]
48. Ognyanova K, Lazer D, Robertson RE, Wilson C. Misinformation in action: fake news exposureis linked to lower trust in media, higher trust in government when your side is in power. Harvard Kennedy Sch Misinformation Rev 2020 Jun 2:1. [doi: 10.37016/mr-2020-024]
49. Shahsavari S, Holur P, Wang T, Tangherlini T, Roychowdhury V. Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news. J Comput Soc Sci 2020 Oct 28:1-39 [FREE Full text] [doi: 10.1007/s42001-020-00086-5] [Medline: 33134595]
50. Tangherlini TR, Shahsavari S, Shahbazi B, Ebrahimzadeh E, Roychowdhury V. An automated pipeline for the discovery of conspiracy and conspiracy theory narrative frameworks: Bridgegate, Pizzagate and storytelling on the web. PLoS One 2020;15(6):e0233879 [FREE Full text] [doi: 10.1371/journal.pone.0233879] [Medline: 32544200]
51. Hossain T, Logan IV RL, Ugarte A, Matsubara Y, Young S, Singh S. COVIDLies: detecting COVID-19 misinformation on social media. 2020 Presented at: 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020; December 2020; Online. [doi: 10.18653/v1/2020.nlpcovid19-2.11]
52. twintproject / twint. GitHub. URL: https://github.com/twintproject/twint [accessed 2020-08-27]
53. Search tweets. Twitter Developer. URL: https://developer.twitter.com/en/docs/tweets/search/ [accessed 2020-08-27]
54. Cho SE, Park HW. A qualitative analysis of cross-cultural new media research: SNS use in Asia and the West. Quality Quantity 2012 Jan 10;47(4):2319-2330. [doi: 10.1007/s11135-011-9658-z]
55. Li Y, Wang X, Lin X, Hajli M. Seeking and sharing health information on social media: a net valence model and cross-cultural comparison. Technological Forecasting Soc Change 2018 Jan;126:28-40. [doi: 10.1016/j.techfore.2016.07.021]
56. Ostrowski DA. Using latent Dirichlet allocation for topic modelling in Twitter. In: Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing. 2015 Presented at: IEEE ICSC 2015; February 7-9, 2015; Anaheim, CA. [doi: 10.1109/icosc.2015.7050858]
57. Adiwardana D, Luong MT, So DR, Hall J, Fiedel N, Thoppilan R, et al. Towards a human-like open-domain chatbot. arXiv. Preprint posted online January 27, 2020 [FREE Full text]
58. Park HW, Park S, Chong M. Conversations and medical news frames on Twitter: infodemiological study on COVID-19 in South Korea. J Med Internet Res 2020 May 05;22(5):e18897 [FREE Full text] [doi: 10.2196/18897] [Medline: 32325426]
59. Shin M, Han S, Park S, Cha M. A risk communication event detection model via contrastive learning. 2020 Presented at: 3rd NLP4IF Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda; December 2020; Barcelona, Spain.
60. Yin J, Wang J. A dirichlet multinomial mixture model-based approach for short text clustering. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014 Presented at: KDD '14; August 2014; New York, NY p. 233-242. [doi: 10.1145/2623330.2623715]
61. Zhao WX, Jiang J, Weng J, He J, Lim EP, Yan H, et al. Comparing Twitter and traditional media using topic models. In: Clough P, Foley C, Gurrin C, Jones GJF, Kraaij W, Lee H, et al, editors. Advances in Information Retrieval 33rd European Conference on IR Research, ECIR 2011, Dublin, Ireland, April 18-21, 2011. Proceedings. Berlin, Heidelberg: Springer; 2011:338-349.
62. Weng J, Lim EP, Jiang J, He Q. TwitterRank: finding topic-sensitive influential twitterers. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining. 2010 Presented at: WSDM '10; February 2010; New York, NY p. 261-270. [doi: 10.1145/1718487.1718520]
63. Papakyriakopoulos O, Medina Serrano JC, Hegelich S. The spread of COVID-19 conspiracy theories on social media and the effect of content moderation. Harvard Kennedy Sch Misinformation Rev 2020 Aug 17:1. [doi: 10.37016/mr-2020-034]
64. Krause NM, Freiling I, Beets B, Brossard D. Fact-checking as risk communication: the multi-layered risk of misinformation in times of COVID-19. J Risk Res 2020 Apr 22;23(7-8):1052-1059. [doi: 10.1080/13669877.2020.1756385]
65. Kim H, Walker D. Leveraging volunteer fact checking to identify misinformation about COVID-19 in social media. Harvard Kennedy Sch Misinformation Rev 2020 May 18:1. [doi: 10.37016/mr-2020-021]

Abbreviations

GT: ground truth
LDA: latent Dirichlet allocation
RQ: research question
SARS: severe acute respiratory syndrome

Edited by C Basch; submitted 27.08.20; peer-reviewed by W Ahmed, T Tangherlini, A Chang; comments to author 21.09.20; revised version received 20.12.20; accepted 03.03.21; published 16.03.21.

Please cite as:
Park S, Han S, Kim J, Molaie MM, Vu HD, Singh K, Han J, Lee W, Cha M
COVID-19 Discourse on Twitter in Four Asian Countries: Case Study of Risk Communication
J Med Internet Res 2021;23(3):e23272
URL: https://www.jmir.org/2021/3/e23272
doi:10.2196/23272
PMID:33684054

Original Paper

Social Media and Emotional Burnout Regulation During the COVID-19 Pandemic: Multilevel Approach

Ruosi Shao[1], MA; Zhen Shi[2], MSc; Di Zhang[2], PhD

[1]The Donald P Bellisairo College of Communications, The Pennsylvania State University, University Park, PA, United States

[2]The Research Center for Journalism and Social Development, Renmin University of China, Beijing, China

Corresponding Author:
Di Zhang, PhD
The Research Center for Journalism and Social Development
Renmin University of China
RM713 Mingde Journalism Building
59 Zhongguancun Rd, Haidian Dist
Beijing, 100872
China
Phone: 86 13801247280
Email: zhangdi204@gmail.com

Abstract

Background: In February 2020, the Chinese government imposed a complete lockdown of Wuhan and other cities in Hubei Province to contain a spike of COVID-19 cases. Although such measures are effective in preventing the spread of the virus, medical professionals strongly voiced a caveat concerning the pandemic emotional burnout at the individual level. Although the lockdown limited individuals' interpersonal communication with people in their social networks, it is common that individuals turn to social media to seek and share health information, exchange social support, and express pandemic-generated feelings.

Objective: Based on a holistic and multilevel perspective, this study examines how pandemic-related emotional exhaustion enacts intrapersonal, interpersonal, and hyperpersonal emotional regulation strategies, and then evaluates the effectiveness of these strategies, with a particular interest in understanding the role of hyperpersonal-level regulation or social media–based regulation.

Methods: Using an online panel, this study sampled 538 Chinese internet users from Hubei Province, the epicenter of the COVID-19 outbreak in China. Survey data collection lasted for 12 days from February 7-18, 2020, two weeks after Hubei Province was placed under quarantine. The sample had an average age of 35 (SD 10.65, range 18-78) years, and a majority were married (n=369, 68.6%).

Results: Using structural equation modeling, this study found that intrapersonal-level (B=0.22; β=.24; *P*<.001) and interpersonal-level (B=0.35; β=.49; *P*<.001) emotional regulation strategies were positively associated with individuals' outcome reappraisal. In contrast with intrapersonal and interpersonal regulations, hyperpersonal (social media–based) regulation strategies, such as disclosing and retweeting negative emotions, were negatively related to the outcome reappraisal (B=–1.00; β=–.80; *P*<.001).

Conclusions: Consistent with previous literature, intrapersonal-level regulation (eg, cognitive reappraisal, mindfulness, and self-kindness) and interpersonal-level supportive interaction may generate a buffering effect on emotional exhaustion and promote individuals' reappraisal toward the stressful situation. However, hyperpersonal-level regulation may exacerbate the experienced negative emotions and impede reappraisal of the pandemic situation. It is speculated that retweeting content that contains pandemic-related stress and anxiety may cause a digital emotion contagion. Individuals who share other people's negative emotional expressions on social media are likely to be affected by the negative affect contagion. More importantly, the possible benefits of intrapersonal and interpersonal emotion regulations may be counteracted by social media or hyperpersonal regulation. This suggests the necessity to conduct social media–based health communication interventions to mitigate the social media–wide negative affect contagion if lockdown policies related to highly infectious diseases are initiated.

(J Med Internet Res 2021;23(3):e27015) doi:10.2196/27015

KEYWORDS

COVID-19; pandemic; emotion regulation; emotional exhaustion; multilevel approach; well-being; emotion; mental health; social media; perspective; strategy; effective; modeling; buffer

Introduction

Background

Since the index case of COVID-19 was first documented in Wuhan on December 1, 2019 [1], the COVID-19 pandemic has swept across the world. On January 23, 2020, the Chinese government imposed a complete lockdown of Wuhan and other cities in Hubei Province, the epicenter of the COVID-19 outbreak in China, to contain the spike in new cases. In an effort to *flatten the curve*, this measure stipulated that only one person from each household was permitted to go outside the home for provisions once every 2 days, except for medical reasons or employment at food stores, pharmacies, or hospitals. Following the order of complete social distancing, local residents in Hubei Province had to drastically modify their lifestyles. This was not only in Hubei, and this can now be observed in other parts of the world such as London. Although such measures are effective in containing the spread of the virus at the societal level, medical professionals strongly voiced a caveat concerning the pandemic emotional burnout at the individual level [2]. The pandemic itself, as well as the drastic social distancing policies, generates unprecedented stressors including perceived severe threats to personal safety, intense fears, strong feelings of being out of control, exhaustion, and loneliness, thus exerting compounding effects on individuals mentally, emotionally, and physically [3,4].

Given the affective cost of coping with the pandemic, health authorities worldwide, including the US Centers for Disease Control and Prevention, have called for more attention to daily emotion regulation among individuals [5]. According to the literature on social psychology and interpersonal communication, traditional emotion regulation strategies include cognitive reappraisal, self-kindness regulation, and interaction with one's support network. In addition to these traditional adjustment strategies, social media has become an important channel for disclosing and regulating negative emotions [6,7]. During the lockdown, individuals in isolation turn to social media to seek and share health information, exchange social support, and express pandemic-generated feelings [8]. Although previous studies suggest that social media use can promote psychological well-being through satisfying individual's need for belonging, receiving informational and emotional support, and reducing stress and loneliness [9], empirical evidence equally reveals that, during the COVID-19 pandemic, social media use exacerbates the anxiety, stress, and depression symptoms people have already experienced [8,10]. The effectiveness of social media–based emotion regulation, along with its relationship with the traditional regulation strategies, calls for a careful and systematic examination.

Additionally, existing COVID-19–related studies on the association between social media use and mental health focus on social media use alone without considering other possible methods of emotion regulation [8,10]. However, the extant literature on emotion regulation and well-being suggests that, although regulation effectiveness differs across each strategy, there is no single best and cure-all solution [7]. Instead, researchers suggest that a combination of several strategies and a certain level of flexibility in choosing regulation strategies predict more desirable outcomes [7]. Thus, researchers have called for adopting a multilevel approach to understanding the emotion regulation process within and across levels, along with the dynamics between regulation strategies across levels [11]. Based on a holistic and multilevel perspective, this study first examined how pandemic-related emotional exhaustion enacts regulation behaviors at each level and then evaluated the effectiveness of these regulation strategies, with a particular interest in depicting the interconnectedness among intrapersonal, interpersonal, and hyperpersonal regulation processes. Therefore, this study can present a more comprehensive picture of emotional regulation strategies during the COVID-19 lockdown than other studies.

This study also has potential practical implications for psychological intervention during lockdowns. The research is based on a survey of respondents from China's Hubei Province, which was collected in early February 2020. During that period, people including experts knew little about the new coronavirus, and Hubei Province implemented perhaps the most stringent but effective COVID-19 lockdown measures in the world. An analysis of an extreme case of scientific uncertainty and emotional exhaustion can offer valuable practical experiences for professionals to design their psychological intervention schemes for similar situations in the future.

Literature Review

Emotional Exhaustion in the COVID-19 Pandemic

Emotional exhaustion has been highlighted as one of the most prominent stress reactions signifying a state of feeling emotionally drained and a depletion of affective resources [12-14]. Exhaustion concerns have been examined in medical, educational, and organizational contexts, where individuals cope with stress and uncertainty on a routine basis [15-17]. The conceptualization of emotional exhaustion addresses its nature of being intensely affective and energy depleting, and therefore directly influences individuals' appraisal and coping with the stressors.

Epidemic studies suggest that infectious diseases characterized by long duration, extra complexity, ambiguity, and excessive demands on coping resources could result in persistent psychopathological consequences such as acute distress, anxiety, and emotional exhaustion [18]. The results from two cross-sectional studies conducted in China and Italy show that nurses and medical professionals participating in COVID-19 emergency-related work experienced moderate to severe emotional exhaustion [19,20]. In addition to medical professionals, the public is also impacted by acute prevalent stressors during the outbreaks of infectious diseases. For

example, Chinese college students reported significant emotional distress and psychological burnout in response to the severe acute respiratory syndrome (SARS) epidemic in 2003 [21,22], and similar emotional distress (eg, anxiety, stress, fear, and depression symptoms) was observed in various countries during the recent COVID-19 pandemic [10,23,24].

When experiencing emotional exhaustion, individuals initiate regulation processes to manage stressor-generated feelings and cope with the situation [25]. Empirical evidence in the past two decades suggests using various regulation strategies such as mindfulness, cognitive restricting, avoidance, and support seeking to facilitate adaptive stress appraisal and predict decreased emotional exhaustion [15,22]. Moreover, the results demonstrate that emotion regulation strategies serve as a buffer against the negative impacts of anxiety, fear, or stress and promote better psychological adjustment to the situation [7].

Reappraisal of the Pandemic Situation Through Multilevel Emotion Regulation

The goal to influence one's emotional trajectory activates the emotion regulation process [26]. According to the transactional theory of stress and coping, individuals constantly appraise external stimuli within their environment, among which those appraised as threatening, harmful, or challenging generate negative emotions, further initiating the regulation process [25]. In the regulation processes, individuals enact various strategies, deliberately or unconsciously, to change the current affective state.

According to the transactional theory, the regulation process involves "constantly changing cognitive and behavioral efforts to manage external and/or internal demands that are appraised as taxing or exceeding the resources of a person" [25]. Individuals can either make efforts to manage the stressor or regulate resultant emotions directly. Therefore, regulation is conceptualized as a dynamic and adaptive process that restores the equilibrium between individuals' appraisal of the environment and one's coping resources. The transactional theory suggests assessing the effectiveness of the regulation process through cognitive reappraisal, including a re-evaluation of the situation's nature (eg, from stressful to benign) and individuals' coping ability (eg, from inadequate to adequate) toward it [27].

Intrapersonal-Level Emotional Regulation

Research on emotion regulation has identified various regulation strategies that vary in terms of their primary impacts on the emotion generative process. Empirical evidence suggests that individuals differ systematically in their use of regulation strategies and the reported effectiveness [28]. In the process model of emotion regulation, Gross [29] proposes to use cognitive reappraisal as a common emotion regulation strategy. Cognitive reappraisal is an antecedent-focused strategy that occurs before the emotion responses have been entirely generated. By reconstructing the emotion-eliciting situation differently, reappraisal alters its emotional impact and the subsequent emotional trajectory. Empirical evidence has demonstrated the effectiveness of reappraisal on downregulating negative emotions, improving environmental mastery, and promoting psychological well-being [28,30].

In recent decades, mindfulness has received more attention in cognitive and behavioral therapy, and has been recommended as an alternative to reappraisal regulation. Mindfulness regulation requires purposefully and nonjudgmentally focusing on the present moment and therefore promotes awareness of one's current state and an openness to accept it [31,32]. The results from a daily diary study reveal that mindfulness regulation predicts emotional well-being, including a lower-level negative affect and a higher-level positive affect [33].

In addition to mindfulness regulation, researchers in positive psychology also suggest the use of self-kindness as an alternative to regulate negative affect. Self-kindness regulation entails understanding, gentleness, and love toward oneself in times of struggle; it not only is teachable through weeklong interventions but also demonstrates the buffering effects on stress and self-criticism [34,35].

Reappraisal, mindfulness, and self-kindness, although occurring at different stages during the emotion-generative process and each with a unique locus of focus, happen all at the intrapersonal level within an individual's cognitive processing. These three regulation strategies are selected to reflect the characteristics of some commonly adopted intrapersonal-level regulations, rather than to provide a comprehensive review of emotion regulation. By involving these three strategies, this study first re-evaluates the extent that intrapersonal-level emotion regulation is associated with emotional exhaustion and outcome reappraisal in the context of the COVID-19 pandemic. In particular, we formed the following hypotheses (H):

- H1a: Pandemic-related emotional exhaustion will be positively associated with intrapersonal-level emotion regulation.
- H2a: Intrapersonal-level regulation will be positively associated with outcome reappraisal of the pandemic situation.

Interpersonal-Level Emotion Regulation and Support Seeking

Support seeking is commonly viewed as an active regulation behavior to cope with stressful situations and emotional distress [36]. It differs from the intrapersonal-level regulation process in its fundamentally social nature. As the "first act" in the process of supportive communication, support seeking is defined as "an intentional communicative activity to elicit supportive actions from others" [37]. When facing health-related stress and anxiety, individuals turn to their friends and family for emotional support to regain coping efficacy toward the situation. A cross-sectional study demonstrated that individuals who engaged in active support-seeking behaviors reported less depressive and anxiety symptoms when coping with SARS-related stressors [22].

The way people seek support from others varies in terms of the directness of one's communication. A direct seeking behavior involves explicitly asking for assistance and clearly discussing the problem or distress [38]. In contrast, indirect seeking is often less informative about one's problem and support needs, and

often expresses one's desire through nonverbal cues such as sighing, fidgeting, and avoiding eye contact [38,39]. The choice of direct and indirect seeking strategies is influenced by various factors including help seekers' characteristics (eg, gender or social skills), appraisal of the current problem (eg, perceived stigma, attribution of the cause, or ego-relevance), and the relational nature with the support provider [36]. Although differing in their effectiveness, both support-seeking strategies are commonly adopted in the interpersonal-level regulation process and have demonstrated effects on improving emotional distress [40]. Thus, we hypothesize the following:

- H1b: Pandemic-related emotional exhaustion will be positively associated with interpersonal-level emotion regulation.
- H2b: Interpersonal-level emotion regulation will be positively associated with outcome reappraisal of the pandemic situation.

Hyperpersonal-Level Emotion Regulation

The prevalence of social technologies has expanded and transformed how individuals share and regulate emotional experiences. Recent studies have witnessed an increase in social media use and pandemic-related online expression, particularly aiming to manage the stress, anxiety, and emotional burnout toward COVID-19 [10,23]. During lockdown and social distancing, individuals express emotional experiences and release pent-up feelings to their social network and beyond through posting on social media [8]. Empirical evidence has shown that individuals who experience a higher intensity of stress provide more frequent and intimate self-disclosure on social media [41]. Therefore, research on emotion regulation proposes adopting online self-disclosure as an alternative regulation strategy and expects it to increase sharers' ability to comprehend and further vent the negative affect [42].

However, the social context of the computer-mediated communication environment fundamentally changes the nature and the psychological outcomes of emotional disclosure. Unlike writing down feelings in a private diary, self-disclosure on social media always involves the presence of the *imaginary audience* who may or may not participate in concrete conversations. Therefore, social media constructs a hyperpersonal communication environment where social context cues such as the appearance and facial expressions of the conversational partner are constantly unavailable, and individuals perceive the possibility of others "reading" and "interacting" with their posts through an overattribution of "the abstract others" being present in the mediated environment [43].

Under the observation of the *imaginary audience*, sharing negative emotions, either through posting one's feelings or retweeting emotional content that reflects one's current affective state, may be regarded as *inappropriate* and could result in psychological maladjustment. The feeling of "inappropriateness" to negative disclosure feelings comes from both the "positivity norm" of internet culture and individuals' self-presentation desires [44,45]. In particular, negative sharing receives less feedback, reduces perceived connectedness in the online community, and generates concerns about the social cost of a negative self-presentation. In line with the theoretical proposition, the results from a daily diary study reveal that disclosing negative emotions on social media increases negative affect, rather than neutralizing it [6].

In addition to posting feelings online, social media provides an alternative way to disclose one's affective state, through retweeting emotional content. Retweeting emotional content, as another type of emotional disclosure, differs from the posting behavior by creating a connection between two individuals, the one who writes the original tweet and the one who shares it. Through retweeting, emotional experiences can be transmitted from one person to another and become *contagious*. Given that the COVID-19 outbreak has generated impacts on a large scale, individuals are more susceptible to take on other people's stress, anxiety, and emotional burnout on social media [10]. Considering the effects of social media self-disclosure and emotion contagion, we further hypothesize that hyperpersonal-level emotion regulation, such as sharing feelings or retweeting emotional content on social media, may generate maladaptive effects.

- H1c: Pandemic-related emotional exhaustion will be positively associated with hyperpersonal-level emotion regulation.
- H2c: Hyperpersonal-level emotion regulation will be negatively associated with outcome reappraisal of the pandemic situation.

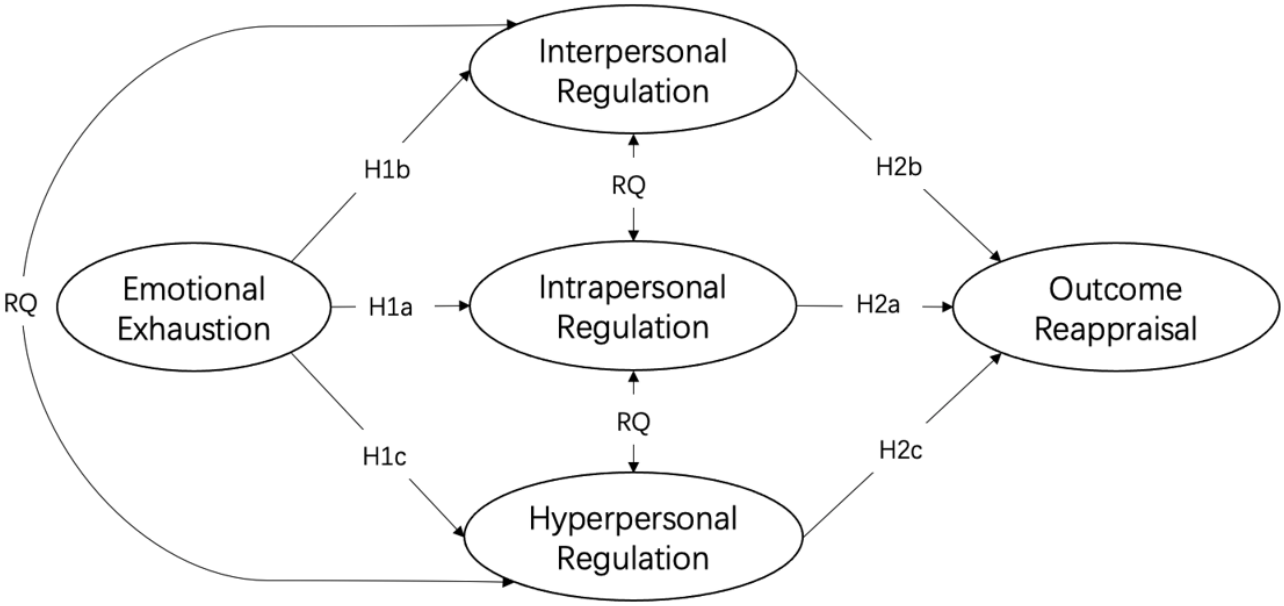
A Multilevel Approach to Emotion Regulation

The process of emotion regulation involves using internal, relational, and societal resources to manage individuals' affective state and, thereby, is fundamentally a multilevel construct. Individuals dynamically adjust regulation strategies across contexts and use emotion regulation within and across levels to maximally succeed in pursuing their own idiosyncratic goals. A growing understanding suggests matching regulation strategies to environmental circumstances and evaluating the effectiveness of emotion regulation holistically, rather than focusing on individual strategies [46].

In this study, we adopt a multilevel approach to depict the interconnectedness between regulation processes at intrapersonal, interpersonal, and hyperpersonal levels [11]. By incorporating a multilevel focus, we examine the system feature of emotion regulation and thereby reveal the mechanism of regulation effectiveness from a broader sense (see Figure 1 for the conceptual model). For example, when individuals engage in emotion regulation at one level, will they simultaneously use regulation strategies at other levels? Furthermore, will regulation processes at different levels generate cumulative effects or counteract each other on coping effectiveness? Thus, we came up with a research question (RQ) about the relationship between regulation processes at three levels.

- RQ: How are emotion regulation processes at intrapersonal, interpersonal, and hyperpersonal levels correlated with each other?

Figure 1. Conceptual model. H: hypothesis; RQ: research question.



Methods

Sample and Procedure

An online survey was conducted to examine individuals' emotional state, regulation strategies, and outcome reappraisal and coping during the COVID-19 outbreak. Data collection lasted for 12 days from February 7 to 18, 2020, two weeks after Hubei Province was placed under quarantine. During the survey period, participants were undergoing quarantine based on government policy and received a link to the online questionnaire.

The survey was first designed in English and then translated into Chinese by two bilingual communication researchers. The final version was further reviewed by the third bilingual researcher with expertise in public health to ensure the accuracy of the translation. Sociocultural background and linguistic nuances were recognized and reflected during the translation. Additionally, we checked the validity of the measurements originally developed in Western contexts. This study measured five latent variables including emotional exhaustion, intrapersonal-level emotion regulation, interpersonal-level emotion regulation, hyperpersonal-level emotion regulation, and outcome reappraisal. All these measurement scales were successfully applied to Chinese respondents in previous studies that demonstrated their validity in the Chinese context [47-51].

Participants (N=538; male: n=273; female: n=265) from China's Hubei Province, where the COVID-19 pandemic first broke out, were recruited from an online survey service in China. After the outbreak in Hubei, the Chinese government implemented strict lockdown and social distancing policies, and most social interactions of people in Hubei took place on mediated channels, which gave the researchers a unique context to understand emotional regulation strategies. The study used quota sampling to draw participants, which is common for online panel research [52]. The quota sampling scheme used age group and gender as the criteria to design subgroups (there were roughly equal numbers of respondents in the six subgroups). The age groups included 18-29 years, 30-39 years, and 40 years or older. The final sample had an average age of 35 (SD 10.65; range 18-78) years and with a majority being married (n=369, 68.6%). The respondents reported very frequent use of social media during the lockdown: for WeChat and microblogs, the average score was 6.01 (SD 1.17) and, for online forums and WeChat Groups, the average scores was 5.47 (SD 1.40; both on a scale of 1-7).

Compared to Chinese census data [53], the sample recruited from the online panel was younger and had slightly more females, and the sample were exclusively internet users. Since the major goal of the study is to examine how emotional regulation through social media is associated with reappraisal, the online panel data can best satisfy our research agenda. The age and gender quota in the final sample is specified to test the proposed conceptual model. Considering that people younger than 35 years account for 82%-83% of China's internet users [54], we oversampled this age group to reflect the majority of internet users in China.

Measures

All measures were on a 7-point scale ranging from 1 (*not at all*) to 7 (*very much*) unless otherwise indicated.

Emotional Exhaustion

Emotional exhaustion was measured with six items adopted from the emotional exhaustion subscale in the Maslach Burnout Inventory–General Survey developed by Schaufeli and colleagues [55]. The Maslach Burnout Inventory is widely used in research to measure emotional exhaustion in various situations and has been validated in studies examining COVID-19–related burnout [19]. We removed three working-related items and kept the remaining six items to measure general emotional exhaustion toward COVID-19, since most people were not back at work when the data was collected. In particular, participants rated to what extent they feel emotionally drained and intense fear and anxiety during the lockdown (mean 20.25, SD 8.64; Cronbach α=.89).

Intrapersonal-Level Emotion Regulation

Three regulation strategies including cognitive reappraisal, self-kindness, and mindfulness were reported to assess intrapersonal-level emotion regulation. Considering that participants were experiencing pandemic-generated stress, we kept the questionnaire brief to increase the completion rate and reduce the response burden. Therefore, when measuring emotion regulation strategies, we only selected items with the highest factor loadings and that were suitable for the research context [56]. Given the one-dimensional nature of each regulation strategy, a short scale (even a single-item one) could demonstrate good validity against equivalent full-scale versions, if not be even preferable [57,58]. In this study, the results of the Cronbach alpha and Spearman–Brown formula, the latter of which was adopted for two-item measurements [59], demonstrated good reliabilities. *Cognitive reappraisal regulation* was measured with five items from the Emotion Regulation Questionnaire by Gross and John [28] asking the level that participants "change the way they think about the pandemic situation when they want to feel less negative emotion" (mean 25.94, SD 5.04; Cronbach α=.76). *Self-kindness regulation* was measured with five items based on the self-kindness subdimension in the Self-Compassion Scale by Neff [60]: "I give myself the caring and tenderness I need during the COVID-19 pandemic" (mean 24.86, SD 5.66; Cronbach α=.81). *Mindfulness regulation* was measured with three items adopted from the mindfulness subdimension in the Self-Compassion Scale by Neff [60]: "When I am feeling down during the pandemic, I try to approach my feelings with openness" (mean 15.54, SD 3.34; Cronbach α=.68).

Interpersonal-Level Emotion Regulation

Two types of support-seeking behaviors (direct and indirect seeking) were assessed based on measurements developed by Derlega and colleagues [36] to evaluate interpersonal-level emotion regulation. *Direct support seeking* was measured with four items (eg, "I tell people the exact emotions I am experiencing because of the pandemic during the supportive interaction" and "I ask people for help when feeling negative about the pandemic situation during supportive interactions"; mean 13.13, SD 6.42; Cronbach α=.89). *Indirect support seeking* was measured with four items (eg, "I sign a lot when talking about the pandemic situation during the supportive interaction"; mean 10.05, SD 5.66; Cronbach α=.85). In addition, *emotional expressivity in support-seeking conversation* was measured with two items from the Emotional Expressivity Scale developed by Kring and colleagues [61] (eg, "I express my feelings about the pandemic in my conversation"; mean 7.82, SD 3.18; $r_{\text{Spearman-Brown}}$=0.77).

Hyperpersonal-Level Emotion Regulation

Regulation strategies on the hyperpersonal-level focus on two emotional expressive behaviors online, posting and retweeting. *Emotional expressive posting* was measured with two items from the Emotional Expressivity Scale developed by Kring and colleagues [61] (eg, "I express my feelings about the pandemic in my online post"; mean 7.24, SD 3.27; $r_{\text{Spearman-Brown}}$=0.79). *Emotional expressive retweeting* was measured by asking respondents to what extent their retweeting content conveys "fear," "anxiety," and "stress" about the pandemic situation (mean 8.22, SD 4.53; Cronbach α=.85).

Outcome Reappraisal

Two types of outcome reappraisal were examined as dependent variables with measures from Holmstrom and Kim [62]. *Emotion-focused reappraisal* was measured with four items asking to what extent respondents, for example, "feel less stressful towards the pandemic situation." *Problem-focused reappraisal* was measured with five items such as "I know more about how to cope with the pandemic situation." An additive index of nine items was created (mean 47.88, SD 9.31; Cronbach α=.89).

Results

Analysis Strategy

Structural equation modeling was used to test the relationship between all components in the proposed model simultaneously while accounting for measurement error. We first ran a confirmatory factor analysis to assess the construct validity of the measurement model and then created a path model using the Lavaan package in R (R Foundation for Statistical Computing). We constructed and examined a five-factor model under maximum likelihood estimation, as the multivariate normality assumption was not violated. A correlation matrix of latent variables (composite indexes were computed) is shown in Table 1.

Table 1. Bivariate correlations between composite indexes of latent variables (N=538).

Variables	1	2	3	4	5	6	7	8	9
1. Outcome reappraisal									
r	—[a]								
P value	—								
2. Emotional exhaustion									
r	–0.35	—							
P value	<.001	—							
3. Self-kindness									
r	0.16	0.15	—						
P value	<.001	.001	—						
4. Mindfulness									
r	0.15	0.12	0.67	—					
P value	<.001	.005	<.001	—					
5. Reappraisal									
r	0.32	0.03	0.61	0.58	—				
P value	<.001	.47	<.001	<.001	—				
6. Direct seeking									
r	–0.06	0.23	0.24	0.21	0.12	—			
P value	.20	<.001	<.001	<.001	.007	—			
7. Indirect seeking									
r	–0.15	0.25	0.15	0.10	0.01	0.71	—		
P value	.001	<.001	<.001	.02	.84	<.001	—		
8. Expressive talk									
r	0.00	0.18	0.12	0.11	0.10	0.44	0.35	—	
P value	.94	<.001	.004	.009	.02	<.001	<.001	—	
9. Expressive posting									
r	0.00	0.23	0.22	0.15	0.09	0.44	0.40	0.50	—
P value	.94	<.001	<.001	.001	.03	<.001	<.001	<.001	—
10. Expressive retweeting									
r	–0.31	0.52	0.05	0.02	–0.08	0.40	0.44	0.29	0.38
P value	<.001	<.001	.23	.65	.07	<.001	<.001	<.001	<.001

[a]Not applicable.

Measurement Model

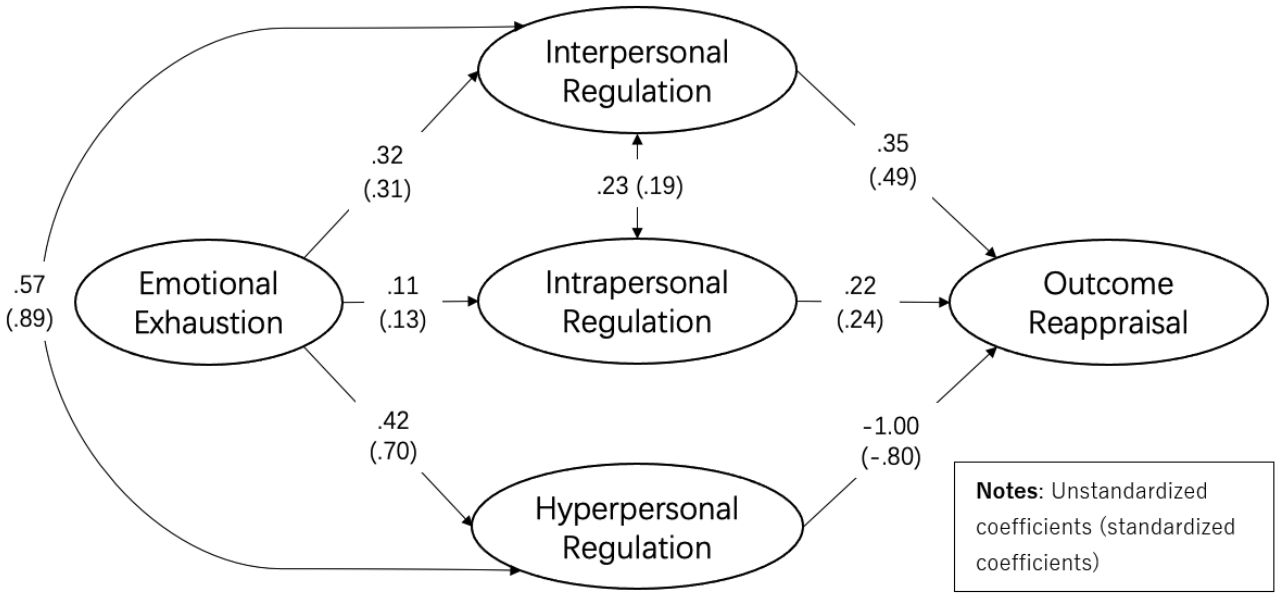
In building the measurement model, the antecedent emotional exhaustion and outcome reappraisal were first-order latent variables. For the three levels of regulation strategies, three second-order latent variables were assessed. The second-order intrapersonal-level emotional regulation included cognitive reappraisal, self-kindness, and mindfulness; interpersonal-level regulation consisted of direct and indirect support seeking and emotional expressivity in support seeking conversation; and hyperpersonal-level regulation encompassed emotional expressive posting and emotional expressive retweeting. The measurement portion of the model demonstrated a good fit (χ^2_{841}/df=2.50; root mean square error of approximation [RMSEA]=0.053; standardized root mean square residual [SRMR]=0.067). A comparative fit index (CFI) was not computed to assess the model fit in this study since, according to statistician Kenny [63], a CFI is not informative when the RMSEA of the null model is less than 0.158 (in this case, it was 0.155).

Path Model and Hypothesis Testing

Further, we turned to the path model to examine the relationships among emotional exhaustion, three levels of emotion regulation, and outcome reappraisal. In the path model (Figure 2), emotional exhaustion was included as the exogenous variable and was left in the model to be associated with three levels of regulation strategies. Intrapersonal, interpersonal, and hyperpersonal

emotion regulation were defined as three second-order latent factors in the model. Each of the three was allowed to be related to outcome reappraisal. To explore the relationships between three levels of the regulation process, we correlated the three regulation strategies (second-order latent variables in the middle of Figure 2). The hypothesized model demonstrated good fit (χ^2_{842}/df=2.68; RMSEA=0.056; SRMR=0.068). Figure 2 displays the path estimates that are statistically significant at the *P* level of .05. The following paragraphs show the results of the hypothesis testing.

Figure 2. Revised model.



H1a hypothesized a positive association between emotional exhaustion and intrapersonal-level emotion regulation. According to the results of the test, H1a was supported (B=0.11; β=.13; *P*=.01). H2a predicted that intrapersonal-level regulation is positively related to outcome reappraisal. The results suggest that H2a was supported (B=0.22; β=.24; *P*<.001).

H1b hypothesized that emotional exhaustion is positively associated with interpersonal-level emotion regulation. H1b was supported (B=0.32; β=.31; *P*<.001). H2b predicted that interpersonal-level emotion regulation has a positive association with outcome reappraisal. The results of the path model suggest that H2b was supported **(**B=0.35; β=.49; *P*<.001).

H1c predicted that emotional exhaustion is positively related to hyperpersonal-level emotion regulation, which was supported (B=0.42; β=.70; *P*<.001). H2c hypothesized a negative relationship between hyperpersonal-level emotion regulation and outcome reappraisal. The results were in line with our prediction (B=–1.00; β=–.80; *P*<.001).

The only research question in this study explores the relationships among the three levels of the regulation process. The results revealed that interpersonal-level emotional regulation is positively associated with intrapersonal-level regulation (B=0.23; β=.19; *P*=.001) and hyperpersonal-level regulation (B=0.57; β=.89; *P*<.001). In contrast, there exists no correlation between intrapersonal-level regulation and hyperpersonal-level regulation (B=0.003; β=.006; *P*=.94).

Discussion

Principal Findings

The COVID-19 pandemic has incurred unprecedented stressors, including the absence of vaccines, loneliness during social distancing, and concerns about the well-being of oneself and loved ones [2]. Given the evidence of affective cost among individuals across different cultures [2], it is crucial to understand and provide health interventions to individuals' emotion regulation processes [8,23]. This study shows that residents in Hubei Province experienced a moderate level of emotional exhaustion during the COVID-19 outbreak and the ensuing unprecedented lockdown. The results find that individuals actively adopted cross-level regulation strategies to cope with the pandemic-generated burnout.

Understanding and Intervening Emotion Regulation Systematically

As the epicenter of the COVID-19 outbreak, Hubei Province implemented stringent social distancing measures and mobility restrictions, such as residents-only entry policies in residential communities, compulsory wearing of face masks, and closing nonessential community services. To Hubei residents that had a lack of understanding toward the pathogenesis of the virus and the medical treatment and the vaccines that were under testing, the uncertainty and anxiety toward the pandemic accumulated to a high level. Under the unique circumstance, this study examined regulation processes among Hubei residents during the lockdown at intrapersonal, interpersonal, and hyperpersonal levels. Consistent with the previous studies [22,30-35], the results showed that intrapersonal-level regulation (eg, cognitive reappraisal, mindfulness, and self-kindness) and

interpersonal-level supportive interaction seem to generate a buffering effect on emotional exhaustion and promote individuals' reappraisal toward the stressful situation. However, according to the results of the study, hyperpersonal-level regulation exacerbates the experienced negative emotions and impedes reappraisal of the pandemic situation.

Based on a multilevel approach recommended by Cook and colleagues [11], this study examined three levels of emotion regulation processes simultaneously and depicted the interconnectedness between levels of regulation behaviors. The results reveal that the adoption of interpersonal-level regulation is commonly combined with intrapersonal or hyperpersonal regulation strategies. Individuals who either regulate one's negative emotions online or internally also turn to friends and family for emotional support to cope with pandemic-generated affect. The complementary nature between interpersonal-level regulation and the other two levels points out the unique contribution of social interactions and the necessity of a certain degree of sociality in the regulation process.

In addition, the positive correlation between intra- and interpersonal level regulation behaviors raises concerns about the disparity regarding regulation resources and ability among individuals. In particular, individuals who are better at internal regulation (eg, cognitive reappraisal, self-kindness, and mindfulness) also engage more frequently in supportive interactions with one's network, leading to a potential gap in emotional well-being between "skillful" regulator with rich support resources and the others who are bad at managing negative emotions and have fewer support resources. Thus, to eliminate the gap, community-wide interventions could prioritize those with fewer network resources and worse regulation ability to promote their emotional well-being.

The Dark Side of Social Media–Based Emotion Regulation

In contrast with intrapersonal and interpersonal emotional regulations, social media–based regulation strategies, such as disclosing and retweeting negative emotions, generate maladaptive effects. The results in this study reveal that individuals who frequently disclose pandemic-related feelings and retweet COVID-19–related negative emotions on social media reported less reappraisal of the stressful situation. The possible maladaptive effects of social media use during the pandemic is consistent with several recent studies in the context of the current COVID-19 outbreak [8,10,64]. Based on the existing literature, this study analyzes social media in a more refined manner. Instead of measuring social media use as a whole, this study enhances our understanding by examining two specific active social media participation behaviors, namely, posting and retweeting. Through posting and retweeting emotional content, individuals use social media as a channel to disclose and regulate their negative emotions. However, the data of this study suggests that disclosing and sharing pandemic-related anxiety and fear on social media cannot help with relieving stress but may lead to a large-scale contagion of emotional burnout.

The possible counter-effects of hyperpersonal regulation may be due to the fact that the pandemic raises societal-level emotional burnout, during which many fellow social media users equally experience excessive demands on coping both emotionally and instrumentally [4]. This may limit their mental resources to provide supportive feedback to negative affective sharing on social media [4]. Therefore, disclosing negative feelings may cause others' fear and anxiety, which further results in self-censorship of one's disclosure behavior as "inappropriate" [44]. In addition, retweeting content that contains pandemic-related stress and anxiety may cause a digital emotion contagion [10]. Individuals who share other people's negative emotional expressions on social media are likely to be affected by the negative affect contagion.

The results of the study sound stern alarms. This study finds that the size of the coefficients for emotional exhaustion's association with intrapersonal regulation is the least, that for interpersonal regulation is mid-sized, and that for hyperpersonal regulation is the largest. In other words, in the era of social media, it is relatively hard for intrapersonal regulation to be activated, while emotional exhaustion can most easily kickoff hyperpersonal and interpersonal regulation. However, it should be noted that the outcomes of the latter two regulation strategies are completely opposite, and the size of the coefficient of hyperpersonal regulation is almost twice that of interpersonal regulation. This means that social media use (B=–1.00; β=–.80; P<.001) may counteract the benefits obtained through interpersonal **(**B=0.35; β=.49; P<.001) and intrapersonal (B=0.22; β=.24; P<.001) regulations. Without proper external interventions, individuals are unlikely to relieve their stress through easily obtained *tools* on their own during the lockdown in a pandemic.

Practical Implications

The results in this study bear practical implications. The survey data was collected right after the outbreak of COVID-19 in Hubei Province, therefore providing a precious opportunity to examine individual's mental well-being and emotion regulation behaviors during the initial stage of the public health risk. According to the crisis and emergency risk communication model, the initial stage of a public emergency is commonly characterized by a high level of uncertainty, a need for reducing stress, and a desire for reassuring self-efficacy [65]. Empirical evidence demonstrates that the high uncertainty and self-relevance nature of the pandemic's initial stage generates moderate to high level emotional exhaustion among individuals. Further, the ability to actively and effectively regulate pandemic-generated emotional exhaustion varies significantly among individuals, and thus calls for public health communication interventions.

The social media–based emotion regulation process generates more impact on outcome reappraisal, compared to intrapersonal and interpersonal regulation. Considering the maladaptive effects of social media–based regulation, it is necessary to conduct effective health communication intervention in the emotional contagion on social media platforms if lockdown policies are to be initiated during public health crises. In the first step, public health agencies may consider using natural language processing techniques to conduct sentiment analysis of social media posts by individuals to closely monitor the stress level and anxiety

tendency, and pinpoint the user segments with such symptoms. Further, these agencies may consider collaborating with social media platforms and using their content recommendation systems to share supportive messages with the target user segments to alleviate their stress and anxiety at the individual level, thus impeding the large-scale contagion of negative affect on social media.

Interpersonal-level regulation is most effective for managing pandemic-generated emotional exhaustion. Considering the benefits of social interactions on coping with pandemic-generated stress (interpersonal-level regulation), public health agencies may consider adopting emerging media technologies such as social robots and augmented and virtual reality to create opportunities for community residents to connect and talk with each other, which social distancing and lockdown policies have made difficult [66]. These new communication technologies could possibly create a mediated environment to fulfill individuals' needs for social interaction and provide an effective channel for community-level intervention.

In addition, intrapersonal-level regulation, though demonstrated as effective and beneficial, can hardly be activated when experiencing emotional exhaustion. Public health interventions should consider focusing on improving awareness of individuals' emotional state and providing guidance on conducting intrapersonal-level regulations. For example, health interventions could use mobile phone–based apps to help individuals practice mindfulness meditation and cognitive reappraisal. Public service advertisements could convey self-kindness messages to the public through broadcasting. Community-level regulation could hold online hangouts to help individuals engage in emotion regulation together virtually.

More importantly, this study implies the limitations of one-to-one psychological counseling. The results suggest that individuals in public health crises attempted various methods to regulate their negative affect. This means that, by incorporating the structural features of emotion regulation, health interventions may adopt a flexible combination of cross-level regulation strategies, based on the contexts in which individuals are embedded.

Limitations and Future Research

The findings from this study should be considered alongside its limitations. A perennial problem with surveys is the bias of retrospective self-reports. Although emotion regulation is commonly defined as a goal-generated activity, some regulation processes may also happen subtly and unconsciously. In addition, the accuracy of recognizing, recalling, and reporting regulation processes varies among individuals based on their emotional intelligence level. Future studies could adopt a daily diary methodology to eliminate potential bias.

Another drawback of the study is the cross-sectional design, and readers should be cautioned about the causality. Future studies could adopt a longitudinal design and test the intervention effectiveness on a 2-week basis. In addition, this study examined individual's emotion regulation behaviors at the initial stage of the COVID-19 outbreak. The initial stage of public health emergencies possesses a unique nature of uncertainty and a need for reassurance. Future studies could extend this study by focusing on other stages and making a systematic comparison.

By adopting a multilevel approach, this study examined emotion regulation processes that take place at different levels. The multilevel approach provides a systematic depiction of regulation processes while allowing for the flexibility of choosing and combining regulation strategies, therefore suiting the theoretical propositions of emotion regulation more appropriately. Future studies should consider adopting the multilevel approach when examining various regulation behaviors.

Acknowledgments

This study was supported by the Ministry of Education Key Project of Key Research Institute of Humanities and Social Sciences (No. 16JJD860008).

Conflicts of Interest

None declared.

References

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
2. Queen D, Harding K. Societal pandemic burnout: a COVID legacy. Int Wound J 2020 Aug;17(4):873-874 [FREE Full text] [doi: 10.1111/iwj.13441] [Medline: 32628363]
3. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). Washington, DC: American Psychiatric Publishing; 2013.
4. Restauri N, Sheridan AD. Burnout and posttraumatic stress disorder in the coronavirus disease 2019 (COVID-19) pandemic: intersection, impact, and interventions. J Am Coll Radiol 2020 Jul;17(7):921-926 [FREE Full text] [doi: 10.1016/j.jacr.2020.05.021] [Medline: 32479798]

5. Healthcare personnel and first responders: how to cope with stress and build resilience during the COVID-19 pandemic. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/mental-health-healthcare.html [accessed 2021-01-01]
6. Choi M, Toma CL. Social sharing through interpersonal media: patterns and effects on emotional well-being. Comput Hum Behav 2014 Jul;36:530-541. [doi: 10.1016/j.chb.2014.04.026]
7. Gross JJ. Emotion regulation: current status and future prospects. Psychological Inquiry 2015 Mar 09;26(1):1-26. [doi: 10.1080/1047840X.2014.940781]
8. Zhong B, Huang Y, Liu Q. Mental health toll from the coronavirus: social media usage reveals Wuhan residents' depression and secondary trauma in the COVID-19 outbreak. Comput Human Behav 2021 Jan;114:106524 [FREE Full text] [doi: 10.1016/j.chb.2020.106524] [Medline: 32836728]
9. Zhan L, Sun Y, Wang N, Zhang X. Understanding the influence of social media on people's life satisfaction through two competing explanatory mechanisms. Aslib J Inf Manage 2016;68(3):347-361. [doi: 10.1108/ajim-12-2015-0195]
10. Prikhidko A, Long H, Wheaton MG. The effect of concerns about COVID-19 on anxiety, stress, parental burnout, and emotion regulation: the role of susceptibility to digital emotion contagion. Front Public Health 2020;8:567250. [doi: 10.3389/fpubh.2020.567250] [Medline: 33392126]
11. Cook JE, Purdie-Vaughns V, Meyer IH, Busch JTA. Intervening within and across levels: a multilevel approach to stigma and public health. Soc Sci Med 2014 Feb;103:101-109. [doi: 10.1016/j.socscimed.2013.09.023] [Medline: 24513229]
12. Figley C. Compassion Fatigue: Coping With Secondary Traumatic Stress Disorder in Those Who Treat the Traumatized. London: Brunner-Routledge; 1995.
13. Halbesleben JRB, Bowler WM. Emotional exhaustion and job performance: the mediating role of motivation. J Appl Psychol 2007 Jan;92(1):93-106. [doi: 10.1037/0021-9010.92.1.93] [Medline: 17227154]
14. Maslach C. Burnout: The Cost of Caring. Englewood Cliffs, NJ: Prentice Hall; 1982.
15. Hülsheger UR, Alberts HJEM, Feinholdt A, Lang JWB. Benefits of mindfulness at work: the role of mindfulness in emotion regulation, emotional exhaustion, and job satisfaction. J Appl Psychol 2013 Mar;98(2):310-325. [doi: 10.1037/a0031313] [Medline: 23276118]
16. Kim B, Lee M, Kim K, Choi H, Lee SM. Longitudinal analysis of academic burnout in Korean middle school students. Stress Health 2015 Oct;31(4):281-289. [doi: 10.1002/smi.2553] [Medline: 24452964]
17. Marjanovic Z, Greenglass ER, Coffey S. The relevance of psychosocial variables and working conditions in predicting nurses' coping strategies during the SARS crisis: an online questionnaire survey. Int J Nurs Stud 2007 Aug;44(6):991-998 [FREE Full text] [doi: 10.1016/j.ijnurstu.2006.02.012] [Medline: 16618485]
18. Tam CWC, Pang EPF, Lam LCW, Chiu HFK. Severe acute respiratory syndrome (SARS) in Hong Kong in 2003: stress and psychological impact among frontline healthcare workers. Psychol Med 2004 Oct;34(7):1197-1204. [doi: 10.1017/s0033291704002247] [Medline: 15697046]
19. Giusti EM, Pedroli E, D'Aniello GE, Stramba Badiale C, Pietrabissa G, Manna C, et al. The psychological impact of the COVID-19 outbreak on health professionals: a cross-sectional study. Front Psychol 2020;11:1684. [doi: 10.3389/fpsyg.2020.01684] [Medline: 32754102]
20. Wan Z, Lian M, Ma H, Cai Z, Xianyu Y. Factors associated with burnout among Chinese nurses during COVID-19 epidemic: a cross-sectional study. BMC Nurs. Preprint posted online June 9, 2020. [FREE Full text] [doi: 10.21203/rs.3.rs-31486/v1]
21. Gan Y, Liu Y, Zhang Y. Flexible coping responses to severe acute respiratory syndrome-related and daily life stressful events. Asian J Soc Psychol 2004;7(1):55-66. [doi: 10.1111/j.1467-839x.2004.00134.x]
22. Main A, Zhou Q, Ma Y, Luecken LJ, Liu X. Relations of SARS-related stressors and coping to Chinese college students' psychological adjustment during the 2003 Beijing SARS epidemic. J Couns Psychol 2011 Jul;58(3):410-423. [doi: 10.1037/a0023632] [Medline: 21574694]
23. Eden AL, Johnson BK, Reinecke L, Grady SM. Media for coping during COVID-19 social distancing: stress, anxiety, and psychological well-being. Front Psychol 2020;11:577639. [doi: 10.3389/fpsyg.2020.577639] [Medline: 33391094]
24. Ye Y, Long T, Liu C, Xu D. The effect of emotion on prosocial tendency: the moderating effect of epidemic severity under the outbreak of COVID-19. Front Psychol 2020;11:588701. [doi: 10.3389/fpsyg.2020.588701] [Medline: 33408665]
25. Lazarus R, Folkman S. Stress, Appraisal, and Coping. New York: Springer; 1984.
26. Gross JJ, Sheppes G, Urry HL. Cognition and Emotion Lecture at the 2010 SPSP Emotion Preconference. Cogn Emot 2011 Aug;25(5):765-781. [doi: 10.1080/02699931.2011.555753] [Medline: 21824019]
27. Biggs A, Brough P, Drummond S. Lazarus and Folkman's psychological stress coping theory. In: Cooper CL, Quick JC, editors. The Handbook of Stress and Health. Oxford, UK: John Wiley & Sons; 2017:349-364.
28. Gross JJ, John OP. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. J Pers Soc Psychol 2003 Aug;85(2):348-362. [doi: 10.1037/0022-3514.85.2.348] [Medline: 12916575]
29. Gross JJ. Emotion regulation in adulthood: timing is everything. Curr Dir Psychol Sci 2016 Jun 23;10(6):214-219. [doi: 10.1111/1467-8721.00152]
30. Ray RD, McRae K, Ochsner KN, Gross JJ. Cognitive reappraisal of negative affect: converging evidence from EMG and self-report. Emotion 2010 Aug;10(4):587-592 [FREE Full text] [doi: 10.1037/a0019015] [Medline: 20677875]

31. Bishop SR. What do we really know about mindfulness-based stress reduction? Psychosom Med 2002;64(1):71-83. [doi: 10.1097/00006842-200201000-00010] [Medline: 11818588]
32. Kabat‐Zinn J. Mindfulness‐based interventions in context: Past, present, and future. Clinical Psychology: Science and Practice 2006 May 11;10(2):144-156. [doi: 10.1093/clipsy.bpg016]
33. Brockman R, Ciarrochi J, Parker P, Kashdan T. Emotion regulation strategies in daily life: mindfulness, cognitive reappraisal and emotion suppression. Cogn Behav Ther 2017 Mar;46(2):91-113. [doi: 10.1080/16506073.2016.1218926] [Medline: 27684649]
34. Neff KD, Kirkpatrick KL, Rude SS. Self-compassion and adaptive psychological functioning. J Res Pers 2007 Feb;41(1):139-154. [doi: 10.1016/j.jrp.2006.03.004]
35. Neff K, Knox M. Self-compassion. In: Zeigler-Hill V, Shackelford TK, editors. Encyclopedia of Personality and Individual Differences. New York: Springer; 2020.
36. Derlega V, Winstead B, Oldfield E, Barbee A. Close relationships and social support in coping with HIV: a test of sensitive interaction systems theory. AIDS Behav 2003 Jun;7(2):119-129. [doi: 10.1023/a:1023990107075] [Medline: 14586197]
37. MacGeorge E, Feng B, Burlseson B. Supportive communication. In: Knapp ML, Daly JA, editors. The SAGE Handbook of Interpersonal Communication. London: Sage; 2011:317-354.
38. Barbee AP, Cunningham MR. An experimental approach to social support communications: interactive coping in close relationships. Ann Int Commun Assoc 2016 May 18;18(1):381-413. [doi: 10.1080/23808985.1995.11678921]
39. Goldsmith DJ. The communicative microdynamics of support. Ann Int Commun Assoc 2016 May 18;18(1):414-433. [doi: 10.1080/23808985.1995.11678922]
40. Jiang L, Drolet A, Kim HS. Age and social support seeking: understanding the role of perceived social costs to others. Pers Soc Psychol Bull 2018 Jul;44(7):1104-1116. [doi: 10.1177/0146167218760798] [Medline: 29552949]
41. Zhang R. The stress-buffering effect of self-disclosure on Facebook: an examination of stressful life events, social support, and mental health among college students. Comput Hum Behav 2017 Oct;75:527-537. [doi: 10.1016/j.chb.2017.05.043]
42. Pennebaker JW. Writing about emotional experiences as a therapeutic process. Psychol Sci 2016 May 06;8(3):162-166. [doi: 10.1111/j.1467-9280.1997.tb00403.x]
43. Walther J, Van DHB, Ramirez A, Burgoon J, Peña J. Interpersonal and hyperpersonal dimensions of computer-mediated communication. In: Sundar SS, editor. The Handbook of the Psychology of Communication Technology. New York: Wiley & Sons; 2015:3-22.
44. Metzler A, Scheithauer H. The long-term benefits of positive self-presentation via profile pictures, number of friends and the initiation of relationships on Facebook for adolescents' self-esteem and the initiation of offline relationships. Front Psychol 2017;8:1981. [doi: 10.3389/fpsyg.2017.01981] [Medline: 29187827]
45. Vogel EA, Rose JP, Crane C. "Transformation Tuesday": temporal context and post valence influence the provision of social support on social media. J Soc Psychol 2018;158(4):446-459. [doi: 10.1080/00224545.2017.1385444] [Medline: 29023225]
46. Bonanno GA, Burton CL. Regulatory flexibility: an individual differences perspective on coping and emotion regulation. Perspect Psychol Sci 2013 Nov;8(6):591-612. [doi: 10.1177/1745691613504116] [Medline: 26173226]
47. Liu W, Chen L, Tu X. Chinese adaptation of Emotion Regulation Questionnaire for Children and Adolescents (ERQ-CCA): a psychometric evaluation in Chinese children. Int J Psychol 2017 Oct;52(5):398-405. [doi: 10.1002/ijop.12233] [Medline: 26611865]
48. Chen R, Sun C, Chen J, Jen H, Kang XL, Kao C, et al. A large-scale survey on trauma, burnout, and posttraumatic growth among nurses during the COVID-19 pandemic. Int J Ment Health Nurs 2021 Feb;30(1):102-116 [FREE Full text] [doi: 10.1111/inm.12796] [Medline: 33107677]
49. Hou J, Wong SY, Lo HH, Mak WW, Ma HS. Validation of a Chinese version of the Five Facet Mindfulness Questionnaire in Hong Kong and development of a short form. Assessment 2014 Jun;21(3):363-371. [doi: 10.1177/1073191113485121] [Medline: 23596271]
50. Lin C, Wu Z, Detels R. Family support, quality of life and concurrent substance use among methadone maintenance therapy clients in China. Public Health 2011 May;125(5):269-274 [FREE Full text] [doi: 10.1016/j.puhe.2011.01.009] [Medline: 21414646]
51. Chan RCK, Shi C, Lui SSY, Ho KKY, Hung KSY, Lam JWS, et al. Validation of the Chinese version of the Clinical Assessment Interview for Negative Symptoms (CAINS): a preliminary report. Front Psychol 2015;6:7. [doi: 10.3389/fpsyg.2015.00007] [Medline: 25698985]
52. Callegaro M, Baker R, Bethlehem J, Göritz A, Krosnick J, Lavrakas P. Online Panel Research: A Data Quality Perspective. London: John Wiley & Sons; 2014.
53. Tabulation on the 2019 population census of the People's Republic of China: Table 2-1 2019. China Statistics Press. 2020 Dec 20. URL: http://www.stats.gov.cn/tjsj/ndsj/2019/indexeh.htm [accessed 2021-01-01]
54. Fang X, Yen DC. Demographics and behavior of internet users in China. Technol Soc 2006 Aug;28(3):363-387. [doi: 10.1016/j.techsoc.2006.06.005]
55. Schaufeli W, Leiter M, Maslach C, Jackson S. The MBI-general survey. In: Maslach C, Jackson SE, Leiter MP, editors. Maslach Burnout Inventory Manual (3rd ed). Palo Alto, CA: Consulting Psychologists Press; 1996:19-26.

56. Raykov T. Alpha if item deleted: a note on loss of criterion validity in scale development if maximizing coefficient alpha. Br J Math Stat Psychol 2008 Nov;61(Pt 2):275-285. [doi: 10.1348/000711007X188520] [Medline: 17535482]
57. Nagy M. Using a single-item approach to measure facet job satisfaction. J Occup Organizational Psychol 2002;75(1):77-86. [doi: 10.1348/096317902167658]
58. Robinson MA. Using multi-item psychometric scales for research and practice in human resource management. Hum Resource Manage 2017 Sep 22;57(3):739-750. [doi: 10.1002/hrm.21852]
59. Eisinga R, Grotenhuis MT, Pelzer B. The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? Int J Public Health 2013 Aug;58(4):637-642. [doi: 10.1007/s00038-012-0416-3] [Medline: 23089674]
60. Neff KD. The development and validation of a scale to measure self-compassion. Self Identity 2003 Jul;2(3):223-250. [doi: 10.1080/15298860309027]
61. Kring AM, Smith DA, Neale JM. Individual differences in dispositional expressiveness: development and validation of the Emotional Expressivity Scale. J Pers Soc Psychol 1994 May;66(5):934-949. [doi: 10.1037//0022-3514.66.5.934] [Medline: 8014836]
62. Holmstrom AJ, Kim S. The mediating role of cognitive reattribution and reappraisal in the esteem support process. Commun Res 2012 Jun 13;42(1):60-86. [doi: 10.1177/0093650212449015]
63. Kenny DA. Measuring model fit. David A Kenny. 2020 Jun 5. URL: http://davidakenny.net/cm/fit.htm [accessed 2021-01-01]
64. Wheaton MG, Prikhidko A, Messner GR. Is fear of COVID-19 contagious? The effects of emotion contagion and social media use on anxiety in response to the coronavirus pandemic. Front Psychol 2020;11:567379. [doi: 10.3389/fpsyg.2020.567379] [Medline: 33469434]
65. Reynolds B, W Seeger M. Crisis and emergency risk communication as an integrative model. J Health Commun 2005;10(1):43-55. [doi: 10.1080/10810730590904571] [Medline: 15764443]
66. Ye J. Pediatric mental and behavioral health in the period of quarantine and social distancing with COVID-19. JMIR Pediatr Parent 2020 Jul 28;3(2):e19867 [FREE Full text] [doi: 10.2196/19867] [Medline: 32634105]

Abbreviations

CFI: comparative fit index
H: hypothesis
MOE: Ministry of Education
RMSEA: root mean square error of approximation
RQ: research question
SARS: severe acute respiratory syndrome
SRMR: standardized root mean square residual

Edited by G Eysenbach, C Basch; submitted 08.01.21; peer-reviewed by H Kim; comments to author 29.01.21; revised version received 03.02.21; accepted 01.03.21; published 16.03.21.

Please cite as:
Shao R, Shi Z, Zhang D
Social Media and Emotional Burnout Regulation During the COVID-19 Pandemic: Multilevel Approach
J Med Internet Res 2021;23(3):e27015
URL: https://www.jmir.org/2021/3/e27015
doi:10.2196/27015
PMID:33661753

Original Paper

Improving Predictions of COVID-19 Preventive Behavior: Development of a Sequential Mediation Model

James A Roberts[1], BA, MBA, PhD; Meredith E David[1], BA, MBA, PhD

Marketing Department, Hankamer School of Business, Baylor University, Waco, TX, United States

Corresponding Author:
James A Roberts, BA, MBA, PhD
Marketing Department
Hankamer School of Business
Baylor University
1 Bear Pl Unit 98007
Waco, TX, 76798-8007
United States
Phone: 1 2547104952
Email: jim_roberts@baylor.edu

Abstract

Background: Since the beginning of the COVID-19 pandemic, social distancing, self-quarantining, wearing masks, and washing hands have become part of the new norm for many, but not all. It appears that such preventive measures are critical to "flattening the curve" of the spread of COVID-19. The public's adoption of such behaviors is an essential component in the battle against what has been referred to as the "invisible enemy."

Objective: The primary objective of this study was to develop a model for predicting COVID-19 preventive behaviors among US college students. The Health Belief Model has a long history of use and empirical support in predicting preventive health behaviors, but it is not without its purported shortcomings. This study identifies a more optimal and defensible combination of variables to explain preventive behaviors among college students. This segment of the US population is critical in helping slow the spread of COVID-19 because of the relative reluctance of college students to perform the needed behaviors given they do not feel susceptible to or fearful of COVID-19.

Methods: For this study, 415 US college students were surveyed via Qualtrics and asked to answer questions regarding their fear of COVID-19, information receptivity (seeking relevant information), perceived knowledge of the disease, self-efficacy, and performance of preventive behaviors. The PROCESS macro (Model 6) was used to test our conceptual model, including predictions involving sequential mediation.

Results: Sequential mediation results show that fear of COVID-19 leads individuals to seek out information regarding the disease, which increases their perceived knowledge and fosters self-efficacy; this is key to driving preventive behaviors.

Conclusions: Self-imposed preventive measures can drastically impact the rate of infection among populations. Based on this study's newly created sequential mediation model, communication strategies for encouraging COVID-19 preventive behaviors are offered. It is clear that college students, and very possibly adults of all ages, must have a healthy fear of COVID-19 to set in motion a process where concerned individuals seek out COVID-19–related information, increasing their store of knowledge concerning the disease, their self-efficacy, and ultimately their likelihood of performing the needed preventive behaviors.

(J Med Internet Res 2021;23(3):e23218) doi:10.2196/23218

KEYWORDS

pandemic; COVID-19; preventive behavior; self-efficacy; prevention; behavior; modeling; student; communication

Introduction

The COVID-19 pandemic is the most recent and devastating of several viral outbreaks that occurred in the past 10-12 years, which include the H1N1 ("swine flu"), severe acute respiratory syndrome (SARS), and Middle East respiratory syndrome (MERS) outbreaks. As of December 10, 2020, there have been 15,040,175 COVID-19 cases diagnosed in the United States and 69,598,919 confirmed cases worldwide. As a result of the COVID-19 pandemic, 285,351 people have died in the United States and 1,582,665 people have died worldwide [1,2]. Social

distancing, self-quarantining, wearing masks, and washing hands have become part of the new norm for many, but not all.

It appears that such preventive measures are critical to "flattening the curve" of the spread of COVID-19. The public's response to prevention messages is an essential component in the battle against what has been referred to as the "invisible enemy" [3]. Effective marketing (preventive) messages that encourage people to social distance, wash their hands, wear masks, and avoid crowds will be critical to ending (or slowing) the spread of COVID-19. However, previous research has shown that increasing awareness of health-related topics is not the same as getting individuals to perform the preventive behaviors needed to avoid being infected and further spreading disease [4].

The Health Belief Model

The Health Belief Model (HBM) had its beginnings in the 1950s. The initial version of the HBM was created by Godfrey Hocbaum, Stephen Kegels, and Irwin Rosenstock in 1952 [5]. At the time, Hocbaum, Kegels, and Rosenstock were working for the US Public Health Service. The original purpose of the HBM was to better understand why people did not partake in public health services such as chest X-rays to screen for tuberculosis. The model was an attempt to integrate Stimulus Response Theory [6] with Social Cognitive Theory (SCT) [7] to better understand various health-related behaviors. Many authors have noted the similarities between the HBM and SCT. SCT, similar to the HBM, is based upon Expectancy-Value Theory (EVT) [8].

SCT stresses the importance of an individual's subjective valuations and his or her expectation that a certain behavior (eg, social distancing during the current pandemic) will lead to the desired end (avoidance of the COVID-19 virus). This combined approach to encouraging healthy behavior is grounded in EVT, based on the early work of Lewin [9] and Vroom [10] and later work by Fishbein and Ajzen [11] and Eccles and several coauthors [8,12-14]. Lewin's theories espoused that it is perceptions of reality, rather than reality itself, that impact behavior [15]. The linchpin of EVT is that incentives or reinforcers do not directly impact health-related behavior but do so by impacting an individual's assessment of the behavior requested and the expectation that said behavior will produce the desired results [8].

The original HBM included four variables: perceived susceptibility (likelihood of contracting the disease), perceived severity (the seriousness/danger of catching the disease), perceived barriers (factors that make taking the required actions difficult), and perceived benefits (positive results of enacting the needed behaviors). In 1988, Rosenstock et al [8] extended the HBM by including a measure of self-efficacy (ie, a person's perception of his or her ability to perform the needed health-related actions). Cues to actions, such as perceived COVID-19 symptoms, personal advice, exposure to media messages, or any other factors that prompt action, are also often included in current applications of the HBM. Over the years, the HBM has been used to explain a wide variety of preventive health behaviors, including during other viral outbreaks (eg, H1N1, SARS, and MERS) similar to the current COVID-19 outbreak [16-18].

Despite the HBM's long history of use and empirical support [8,16,19], it is not without purported shortcomings [20,21]. Several research studies suggest that the percentage of variance explained by the HBM across a variety of preventive measures is low (average R^2 of about 21%) [21] and that the original HBM and its extensions have not used an optimal combination of its most commonly used predictor variables.

Manika and Golden [16] tested an extended version of the HBM. The authors examined the predictive ability of five variables in explaining the likelihood of engaging in preventive behaviors related to the H1N1 pandemic: perceived knowledge, stored knowledge, perceived threat, perceived self-efficacy, and information receptivity (seeking). The researchers entered all five variables simultaneously into a regression model. Of the five independent variables entered into the equation, information receptivity and perceived threat had the strongest coefficients (.39 and .35, respectively). The overall model's R^2 was .46. Later researchers noted that this may not be an optimal combination of predictor variables [20,21]. All the predictor variables were entered into the regression analysis simultaneously.

Orji et al [21] added four new variables to Manika and Golden's [16] extension of the HBM: self-identity, perceived importance, consideration of future consequences, and concern for appearance. The authors also tested all possible relationships/interactions between the original HBM variables. In testing the ability of their extended HBM to predict healthy eating habits, the authors found that the extended model had an R^2 of .71 compared to .40 for the original HBM. One's belief that one can do what is needed to enact behaviors (self-efficacy) was found to be the strongest predictor of preventive behaviors. Results also shed light on several possible combinations of the original HBM variables that could help better explain individual health-seeking behavior.

Jones et al [20] also found that alternative combinations of the HBM's variables can improve its predictive ability. The authors conducted a survey of 1337 Indiana residents after an 8-month flu vaccine campaign that was based on the HBM. Initial findings showed that exposure to the campaign increased the likelihood of getting a flu shot. The study also found an indirect effect of exposure on vaccine behavior through perceived barriers and that perceived threat was moderated by self-efficacy. Perceived barriers and benefits were also part of a serial mediation chain. The authors conclude that the relationships between the variables of the HBM are complex, may help explain earlier inconsistent results, and should be an important focus of future research.

An Improved Model for Predicting COVID-19 Preventive Behavior

The present model (Figure 1) extends the additive model tested by Manika and Golden [16]. Extensions include the following: (1) use of the newly created and validated "fear of COVID-19" measure [22], (2) use of three separate measures of

health-seeking behaviors (including measures developed by Masuri et al [17] and Manika and Golden [16], as well as a brief measure of social distancing created for this study) as the study's dependent variables, and (3) a new combination of predictor variables. We identify and test a sequential mediation model that uses four of the five predictor variables used in the Manika and Golden model. The current model does not include stored knowledge as an independent variable because the Manika and Golden study [16] found that it is not actual but perceived knowledge that impacts preventive health behaviors.

Figure 1. Conceptual model.

Fear of COVID-19 [22] begins the process of encouraging COVID-19 preventive behaviors. As stated by Ahorsu et al [22], "One characteristic nature of infectious disease compared with other conditions is fear." The scale was developed specifically to gauge fear of COVID-19 and was found to be a valid measure. If an individual thinks that they are not susceptible to getting the virus and that the virus is not a serious threat, they are less likely to engage in preventive behaviors [16,20]. A good example is the 2020 college spring break in the United States, where many students congregated on the nation's beaches during the COVID-19 pandemic because they did not feel that COVID-19 was a serious threat [23].

If someone is fearful of the COVID-19 virus, we hypothesize that they are more likely to search for information that can help them avoid contracting the disease. This is the information receptivity variable in the Manika and Golden [16] version of the HBM. Moorman and Matulich [24] tested a similar model to explain preventive health behaviors and found that individuals with higher levels of health motivation were more likely to search for health-related information than less motivated individuals. The authors define health motivation as the arousal level to engage in preventive health behaviors. Items of their health motivation scale measured anxiety or worry about risks to one's health. Hence, fear of COVID-19 will likely increase one's motivation to search for COVID-19–related information.

This increased search for information regarding the COVID-19 virus is hypothesized to lead to a greater level of perceived knowledge. With the emergence of the COVID-19 pandemic, there has been a flood of both information and misinformation about the virus [25]. Depending on the person and media accessed, this increased access to information may or may not lead to appropriate health-related COVID-19 preventive behaviors. Manika and Golden [16] found that perceived knowledge quantity regarding the H1N1 pandemic was positively associated with H1N1 prevention behaviors.

Higher levels of perceived knowledge are hypothesized to lead to greater self-efficacy. Self-efficacy is best understood as an individual's perception that he or she can perform the suggested health-related behaviors. The more confident an individual is that he or she can perform a certain task, the more likely it is that they will engage in the relevant preventive behavior [26]. In Moorman and Matulich's [24] comprehensive review of the preventive health behavior literature, "health ability," a person's perceived ability to perform health behaviors, was found to predict preventive health behaviors. Self-efficacy has been found to be the most important predictor of health-related behaviors [4,21]. Being equipped with perceived knowledge regarding the behaviors needed to avoid contracting COVID-19 and having the confidence that one can perform these behaviors will increase the likelihood of performing COVID-19 preventive behaviors.

To summarize, the sequential mediation model presented and tested in this study hypothesizes that COVID-19 preventive behaviors are driven by a process of sequential mediation in which fear of COVID-19 leads to information receptivity, which in turn leads to perceived knowledge that drives self-efficacy and ultimately results in preventive actions. This proposed sequential mediation model is a potentially significant improvement over previous attempts to predict a variety of health-related behaviors using the HBM or variations thereof. The frequent use of an "additive model" approach to entering independent variables has come under scrutiny and calls have been made to explore other possible causal models [19,27]. This call has been echoed by Champion and Skinner [28], who see the need to better define the relationships between the predictor variables of health-seeking behavior. Jones et al [20] state that establishing a more optimal ordering of such variables will "advance theory and practice by improving evaluation, identifying relative importance of the constructs, and suggesting new postulates for behavior change."

This study's results provide such insights and offer clear directions on what types of preventive messages will be most effective in motivating individuals to incorporate behaviors into their daily routine that help keep them safe from COVID-19. Given the severity of the consequences (both personal and economic) of the COVID-19 pandemic and the importance of individual behaviors in "flattening the curve," research that expands our understanding of what needs to be done to combat this "invisible enemy" is critical.

Methods

Overview

The study design included a cross-sectional survey designed in Qualtrics, which was administered online to participants, who completed the study in one sitting at a time that was convenient for them. The study data were collected using Qualtrics to administer questionnaires to 415 undergraduate students in the

United States (49% male; mean age 21 [SD 5.41] years). Participants were recruited based on being enrolled in a subject pool at the authors' university. Data inclusion criteria were broad such that the subject population included students across multiple majors and class identifications who were aged ≥18 years and participated in the study in exchange for course credit. There were no exclusion criteria and responses were accepted from all students who chose to complete the study. In an effort to minimize any potential response bias, participants were provided with a consent form to complete as part of the study, in which they were informed that their participation is voluntary, they can choose to terminate their participation in the study at any point if they so desire, and they will not be penalized if they choose not to participate in the study.

In addition, informed consent forms were completed by all study participants, in which they were informed that there were no risks involved in participating in the study and that their responses and all data collected would be recorded and entered into a spreadsheet for analysis in a manner whereby all participants will remain anonymous to the researchers, no personal identifying information would be requested, and all data would remain confidential. Most participants were White (n=313, 75%), followed by Hispanic (n=47, 11%), Asian (n=32, 8%), and African American (n=15, 4%).

Measures

Fear of COVID-19

Fear of COVID-19 (α=.89, mean 2.06 [SD 1.87]) was assessed using the 7-item fear of coronavirus-19 scale developed and validated by Ahorsu et al [22]. Example items include "It makes me uncomfortable to think about coronavirus-19," "When watching news and stories about coronavirus-19 on social media, I become nervous or anxious," and "My hands become clammy when I think about coronavirus-19." Response categories range from "strongly disagree" (1) to "strongly agree" (5).

Information Receptivity

Information receptivity (α=.84, mean 4.02 [SD 1.53]) was measured using items adapted from Manika and Golden's [16] 3-item Information Receptivity Scale, which assessed how receptive individuals were to H1N1 information and how actively they sought such information. An example item is "I actively search for information about the H1N1 (swine) flu," which was adapted in this study to "I actively search for information about the coronavirus (COVID-19)." Participants indicated their agreement with the three items using a 7-point scale anchored with strongly disagree and strongly agree.

Perceived Knowledge

Participants' perceived knowledge (α=.84, mean 3.54 [SD .80]) was assessed using an established 4-item measure [16]. Specifically, participants used a 5-point response format (where 1=nothing and 5=a lot) to respond to the following statements: "In general, how much do you think you know about COVID-19," "How much do you think you know about protecting yourself from getting COVID-19," "How much do you think you know about the ways a person can and cannot get COVID-19," and "Please rate your knowledge of COVID-19 compared to the average person."

Perceived Self-Efficacy

Perceived self-efficacy (α=.91, mean 5.56 [SD 1.11]) was assessed using 5 items adapted from the Manika and Golden [16] scale, which was based on Bandura's [7] concept of self-efficacy and assessed how confident individuals felt in their ability to make H1N1 flu prevention choices. An example item is "How confident do you feel about your ability to use your knowledge of H1N1 (swine) flu in making everyday activity choices," which was adapted in this study to "How confident do you feel about your ability to use your knowledge of COVID-19 in making everyday activity choices?" Participants responded to the items on a 7-point scale anchored with not at all confident and completely confident.

Prevention Behavior

The Prevention Behavior Scale by Manika and Golden [16] was used to measure prevention behaviors (α=.85, mean 6.90 [SD 2.17]). The original scale includes the following 4 items: "It is important to me to do everything I reasonably can to avoid getting the H1N1 (swine) flu," "I actively seek information on how I can prevent myself from getting the H1N1 (swine) flu," "I am doing all that I know to do to prevent myself from getting the H1N1 (swine) flu," and "I have changed my behavior to try to avoid getting the H1N1 (swine) flu." For this study, "H1N1 (swine) flu" was changed to "COVID-19." Participants responded to the items using a 7-point Likert scale anchored with strongly disagree and strongly agree. We included two additional measures of prevention behavior that were aimed at more directly assessing changes in behaviors related to COVID-19 prevention. Specifically, we included a measure of actual changes in behaviors, measured using the Masuri et al [17] scale, which is a list of 10 health-related behaviors that participants answer with a "yes" or "no" response. In addition, we included two items that made up the social distancing measure developed for this study (α=.92). These additional measures are reflective of the primary thrust of current preventive message campaigns to social distance to protect oneself and to avoid spreading COVID-19.

Data Analysis

The study data were analyzed in SPSS Statistics (version 25; IBM Corp). Since data for all the study measures were obtained from the same source, common method bias could be a potential issue. Thus, we performed the Lindell and Whitney [29] marker variable procedure to assess whether common method bias was likely to affect the results. Correlations between the marker variable item and each study measure were small (ranging from –.06 to .08) and nonsignificant; thus, it is unlikely that common method bias affected the results [29,30]. The PROCESS macro (Model 6) [31] was used to test our conceptual model, including predictions involving sequential mediation [32]. The Preacher and Hayes [31] PROCESS macro for SPSS was used for several key reasons. This method uses an ordinary least squares path analysis to estimate model coefficients and assess the indirect and/or direct effects of variables in the model [32,33]. In addition, the PROCESS models use a bootstrapping procedure

(n=5000), which does not rely on any assumptions about the normality of the sampling distribution, to calculate the bias-corrected 95% CIs associated with the statistical significance of the indirect effects [31,33,34]. As stated by Li and Tan [35], the Preacher and Hayes [31] PROCESS macro is "specially designed to conduct multiple mediations by including the bootstrapping function while allowing for the estimation of the indirect effect for each mediator. More importantly, this macro accounts for the effects of the control variables that are not easily implemented in structural equation modeling" (for a review of this method, see [31]). Of note, the analyses presented below were also conducted with gender as a control variable; the inclusion of gender did not impact the results presented below.

Results

To begin, the PROCESS macro (Model 6) [31] tests the relationship between fear of COVID-19 and information receptivity. The results ($F_{1,414}$=32.39; P<.01; R^2=.07) indicate that fear of COVID-19 is positively associated with information receptivity (β=.48; P<.001). Next, the model tests whether fear of COVID-19 and information receptivity are directly associated with perceived knowledge. The results ($F_{2,413}$=96.57; P<.01; R^2=.32) indicate that information receptivity is directly associated with knowledge (β=.30; P<.001). However, fear of COVID-19 is not directly associated with perceived knowledge (P=.55). The model next tests the relationship that fear of COVID-19, information receptivity, and knowledge have with perceived self-efficacy. The results ($F_{3,412}$=28.00; P<.001; R^2=.17) show a significant relationship between knowledge and perceived self-efficacy (β=.55; P<.001). Fear of COVID-19 (P=.16) and information receptivity (P=.50) are not directly associated with perceived self-efficacy.

Next, the model tests whether fear of COVID-19, information receptivity, knowledge, and perceived self-efficacy are directly associated with prevention behaviors. The results ($F_{4,411}$=11.99; P<.001; R^2=.11) indicate that perceived self-efficacy (β=.44; P<.001), information receptivity (β=.25; P=.003), and fear of COVID-19 (β=.30; P=.02) are directly associated with prevention behaviors. However, knowledge is not directly associated with prevention behaviors (P=.29). Importantly, the results show support for the predicted sequential mediation (β=.035 [SE .014], 95% CI .012-.064), such that fear of COVID-19 is indirectly associated with prevention behaviors via information receptivity, followed by knowledge, and then perceived self-efficacy. A summary of all direct and indirect paths tested in the model is provided in Table 1.

Table 1. Sequential mediation results including all direct and indirect effects[a].

Paths tested	β coefficient	SE	Lower limit of 95% CI	Upper limit of 95% CI
Fear of COVID-19 → information receptivity	.48	.08	.32	.65
Fear of COVID-19 → perceived knowledge	–.02	.04	–.10	.05
Information receptivity → perceived knowledge	.30	.02	.26	.34
Fear of COVID-19 → perceived self-efficacy	–.09	.06	–.21	.03
Information receptivity → perceived self-efficacy	.03	.04	–.05	.11
Perceived knowledge → perceived self-efficacy	.55	.08	.40	.70
Fear of COVID-19 → prevention behaviors	.30	.12	.06	.55
Information receptivity → prevention behaviors	.25	.08	.09	.41
Perceived knowledge → prevention behaviors	–.17	.16	–.49	.15
Perceived self-efficacy → prevention behaviors	.44	.10	.25	.64
Fear of COVID-19 → information receptivity → prevention behavior	–.12	.04	.04	.21
Fear of COVID-19 → perceived knowledge → prevention behavior	<.01	.01	–.02	.03
Fear of COVID-19 → perceived self-efficacy → prevention behavior	–.04	.04	–.12	.02
Fear of COVID-19 → information receptivity → perceived knowledge → prevention behavior	–.03	.03	–.09	.03
Fear of COVID-19 → information receptivity → perceived self-efficacy → prevention behavior	.01	.01	–.01	.03
Fear of COVID-19 → perceived knowledge → perceived self-efficacy → prevention behavior	–.01	.01	–.03	.02
Fear of COVID-19 → information receptivity → perceived knowledge → perceived self-efficacy → prevention behavior	.04	.01	.01	.06

[a]Results obtained with bootstrapping (n=5000).

Similar results were found using each of the alternative measures of prevention behaviors (alternative measure 1 [prevention

behavior scale]: $F_{4,411}$=90.10; P<.001; R^2=.47; β=.042 [SE .012], 95% CI .021-.069; alternative measure 2 [social distancing]: $F_{4,411}$=39.72; P<.001; R^2=.28; β=.047 [SE .014], 95% CI .023-.077). That is, the results across three different measures of prevention behavior reveal a significant indirect relationship between fear of COVID-19 and preventive behavior.

Discussion

Principal Findings

Self-imposed preventive measures can drastically impact the rate of infection within a population. Public policy officials must consider the economic impact of a pandemic and be highly efficient with the limited resources that can be allocated to marketing communications [36]. While COVID-19 vaccination programs have begun, it is still of utmost importance that the public continues to take preventive measures. Indeed, simple behavioral changes taken by individuals can serve as a highly effective and low-cost method for controlling pandemics [25,37], particularly in the early stages when treatments and vaccinations are not readily available. Our findings suggest that behavioral change is largely dependent on one's confidence in one's own ability to execute behaviors that will reduce the likelihood of contracting or spreading the disease. Importantly, our sequential mediation results show that fear of COVID-19 leads individuals to seek out information regarding the disease, which increases their knowledge and fosters self-efficacy, a key to driving precautionary behavior. Thus, it is vitally important that government and public policy officials communicate the seriousness of the disease and the riskiness of not taking action to reduce the spread of the infection. Specifically, communications should evoke or instill fear among the public, because it is this fear that will enhance information receptivity, knowledge, self-efficacy, and ultimately changes in behavior.

Of note, our results show that fear of COVID-19 itself is not directly associated with preventive behavior. Instead, fear of COVID-19 was only associated with preventive action when individuals sought out knowledge of appropriate preventive measures and felt confident in their ability to take such actions. As stated by Brug et al [38], "risk perceptions as well as efficacy beliefs in the early stages of a possible pandemic are dependent on communications with and between the members of the groups at risk. Risk communication messages that are not comprehended by the public at risk, or communication of conflicting risk messages will result in lack of precautionary actions." The widespread confusion over the efficacy of wearing masks during the current pandemic is one such example.

A system needs to be in place to disseminate accurate information about the virus and appropriate precautionary measures that individuals should take. A system also needs to exist to deter, counteract, and quickly terminate any misinformation that may be spread through news outlets and, more likely, through social media. It is crucial that the information offered is not only accurate but also trustworthy. Thus, consistency is key, and an integrated marketing communications approach would be helpful. Indeed, research has shown that repetitive mentions of preventive actions can serve to reduce fear surrounding a disease, and ultimately increase preventive behaviors, whereas the opposite is true when individuals question the trustworthiness of the information [39].

The overarching objective of this study was to create and test a model for predicting COVID-19 preventive behaviors that addresses the shortcomings of the HBM, and to use this information to formulate more effective health communications to increase the adoption of behaviors that slow the spread of COVID-19. The HBM has a long history of use and empirical support in predicting and explaining a range of health-related behaviors [19]. However, an improved model is needed, given criticisms that note the explanatory abilities of the HBM need to be improved. The resulting sequential mediation model presented herein answers the call for a better definition of the relationships between the antecedents of health-seeking behavior. The present model includes the measure of fear of COVID-19, which adds a validated scale that addresses the COVID-19 pandemic specifically. A criticism of the commonly used scales of the HBM is that they have not been adequately validated [5]. A valid scale to measure the perceived threat and fear of COVID-19 is essential given the important role perceived susceptibility and fear of a disease play in getting individuals to act to lessen their risk of contracting a disease or avoiding other negative health outcomes.

The use of three separate measures of COVID-19 health-seeking behaviors, including items to measure social distancing, is also an important contribution of this study. Items of the Manika and Golden [16] prevention behavior scale were answered on a 7-point Likert scale, while the Masuri et al [17] scale was a list of questions about health-related behaviors that were answered with a "yes" or "no" response. The 2-item social distancing scale added two behaviors (social distancing and quarantining) that were not part of the other two scales but are reflective of the primary thrust of current preventive message campaigns to social distance to protect oneself and to avoid spreading COVID-19.

Data generated from the current survey provide clear directions on how to best encourage COVID-19 preventive behaviors. The present model begins with fear of COVID-19. This is consistent with the HBM model, which states that fear of the disease or another health problem must be present and that individuals must perceive that they are susceptible to the disease or health problem [8]. A good example of the importance of fear of and susceptibility to COVID-19 is evident in young adults, especially college students. Given the general tendency of the young to feel a certain sense of invulnerability to health risks, it has been important to inform them that, although they might not be at high risk for severe disease if they were to contract COVID-19, they could spread the disease to their parents, grandparents, and other older adults. According to Orji et al [21], people of all ages tend to underestimate their likelihood of contracting diseases or experiencing other health problems. Thus, it is important that COVID-19 marketing communications stress the high transmissibility of the virus (susceptibility), as well as the potential severity of the disease (fear).

As suggested by Manika and Golden et al [16], these "fear appeal" messages must be consistent across all media.

Conveying consistent and clear messages of the dangers of COVID-19 is made difficult by social media. Research has shown that people are exposed to considerable amounts of misinformation online, which can undermine the effectiveness of public health campaigns [40,41]. Social media health campaigns are critical to informing the public about proper health behaviors while also dispelling any misinformation that may be being spread online.

This study found that fear of COVID-19 was positively associated with information receptivity. Information receptivity measures how actively an individual searches for COVID-19–related information. Manika and Golden [16] found that people who were more receptive to H1N1 information were also more likely to engage in H1N1 prevention behaviors. The authors concluded that being receptive to information signals an "individual's readiness to act, thus it is more likely that the individual will take appropriate prevention measures." The importance of information receptivity underlines the necessity of having consistent and clear messages—delivered across all media—regarding what types of behaviors are recommended to increase compliance.

In this model, greater information receptivity, not surprisingly, is associated with higher levels of perceived knowledge. The model used perceived knowledge, as opposed to stored (or actual) knowledge, given that research has found that it is one's perceived knowledge (not stored) that leads to requested health-related behaviors [16]. This study also found that perceived knowledge is positively associated with self-efficacy. The more a person believes he or she is equipped with the needed knowledge regarding COVID-19, the more confident they will be that they can perform the behaviors needed to protect themselves from contracting the virus. Health messages must be clear about what types of behaviors are needed and the ease of performing those behaviors.

The public needs to be convinced that behaviors such as washing their hands, wearing masks, social distancing, and self-quarantining will lead to a valued outcome (eg, avoiding contracting COVID-19, reopening the economy). Self-efficacy must be at the forefront of any media campaign. This is particularly true given that self-efficacy is the final variable in our sequential mediation model driving COVID-19 preventive behaviors.

This study used three separate measures of COVID-19 preventive behaviors and found that the 4-item measure developed by Manika and Golden [16] worked the best. The R^2 of .47 suggests that the model does a good job of explaining variation in COVID-19 prevention behavior, but also that a significant percentage of such behavior remains unexplained. It could be, as exemplified by Orji et al [21], that new variables need to be added to the present model. A higher R^2 value may also be achieved by maintaining the current variables of the present model while also including variables already designated in the HBM but not used in this study. Cues to action such as experiencing possible symptoms of COVID-19, conversations among friends and colleagues, and the amount and type of media exposure would all likely increase the explanatory capability of the model. Additionally, a broader array of sociodemographic measures would also increase the R^2 value. Although not found in this study, sex differences have previously been found, with females more anxious about the disease and more likely to perform preventive health behaviors. Age, socioeconomic status, and ethnicity have been identified as possible factors that impact health-related behaviors and should be incorporated into future models.

Limitations and Future Research Directions

Although this survey is the first to apply a sequential mediation model to better understand COVID-19 preventive behaviors, it is not without limitations. First, the sample was from only one segment of the population, albeit an important one. Future research should sample older adults and younger adults, as well as people from specific ethnic populations. In the United States, the African American and Latino populations have been impacted more severely than the White population. A survey of these segments of the population with the present model might provide insights into what encourages or discourages the needed preventive behaviors across ethnic/racial groups. In addition, future research should consider how individuals' backgrounds, medical histories, and other individual difference variables, including those related to public health knowledge, may impact the results found as well as the general performance of the HBM.

A second limitation is the study's correlational nature. Future causal and/or longitudinal research is needed that can test the links in the present model and provide more reliable insights into the direction of the causal flow. A third possible limitation and fodder for future research is the further expansion of the present model. As called for by Jones et al [20], a deeper investigation of cues to action could improve explanatory capacity. Internal cues such as perceived COVID-19 symptoms and external cues such as media exposure (both amount and type) might be better addressed separately. As Jones et al [20] suggest, it might also prove beneficial to investigate manipulated cues to action like public service campaigns, media messages, and interventions, as well as more organic cues to action like illness in the family, discussions among friends and family, news stories, and high profile people contracting the disease.

Conclusion

Drawing upon earlier attempts to explain health-seeking behavior, this study created and tested a sequential mediation model to explain COVID-19 preventive behaviors. The new model's R^2 was .47, which is considerably higher than the average R^2 of approximately .20 across numerous studies using the HBM [21]. The study also identified a better-defined model of the relationships between its predictor variables. Both results were in response to calls to improve the purported shortcomings of the HBM in predicting health-related behavior. Finally, this study's results provide clear directions for creating effective strategies and messages to encourage the behaviors needed to slow, if not eradicate, the most serious pandemic of the past 100 years.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. CDC COVID Data Tracker. URL: https://covid.cdc.gov/covid-data-tracker/#cases_totalcases [accessed 2020-09-16]
2. Worldometers. COVID-19 Coronavirus Pandemic. URL: http://www.worldometers.info/coronavirus/ [accessed 2020-12-10]
3. Cathey L. Trump now calling coronavirus fight a 'war' with an 'invisible enemy'. ABC News. 2020 Mar 17. URL: https://abcnews.go.com/Politics/trump-coronavirus-task-force-economic-%20public-health-steps/story?id=69646672 [accessed 2020-05-11]
4. Glanz K, Rimer BK, Viswanath K. Health Behavior and Health Education: Theory, Research, and Practice. Hoboken, NJ, USA: John Wiley & Sons; 1991.
5. Taylor D, Bury M, Campling N, Carter S, Garfied S, Newbould J, et al. A Review of the use of the Health Belief Model (HBM), the Theory of Reasoned Action (TRA), the Theory of Planned Behavior (TPB) and the Trans-Theoretical Model (TTM) to study and predict health related behavior change. 2007. URL: https://warwick.ac.uk/fac/sci/med/study/ugr/mbchb/phase1_08/semester2/healthpsychology/nice-doh_draft_review_of_health_behaviour_theories.pdf [accessed 2021-03-10]
6. Skinner BF. The Behavior of Organisms. New York, NY, USA: Appleton-Century-Crofts Ltd; 1938.
7. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
8. Rosenstock I, Strecher V, Becker M. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
9. Lewin K. The Nature of Field Theory. In: Marx MH, editor. Psychological Theory. New York, NY, USA: Macmillan; 1951.
10. Vroom VH. Work and Motivation. Hoboken, NJ, USA: Wiley; 1964.
11. Fishbein M, Azjen I. Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Reading, MA, USA: Addison-Wesley; 1975.
12. Eccles J, Adler TF, Futterman R, Goff SB, Kaczala CM, Meece JL. Expectancies, values and academic behaviors. In: Spence JT, editor. Achievement and Achievement Motives. San Francisco, CA, USA: WH Freeman; 1983:75-146.
13. Eccles JS, Wigfield A. Motivational beliefs, values, and goals. Annu Rev Psychol 2002;53:109-132 [FREE Full text] [doi: 10.1146/annurev.psych.53.100901.135153] [Medline: 11752481]
14. Wigfield A, Eccles JS. The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In: Wigfield A, Eccles JS, editors. Development of Achievement Motivation. San Diego, CA, USA: Academic Press; 2001:91-120.
15. Coghlan D, Jacobs C. Kurt Lewin on Reeducation. The Journal of Applied Behavioral Science 2016 Jul 26;41(4):444-457 [FREE Full text] [doi: 10.1177/0021886305277275]
16. Manika D, Golden LL. Self-efficacy, threat, knowledge and information receptivity: Exploring pandemic prevention behaviors to enhance societal welfare. Academy of Health Care Management Journal 2011;7(1):31-44.
17. Masuri MG, Isa K, Tahir M, Hassan H, Hassan Z, Jamhuri FZ, et al. Behaviour Response among Health Sciences Students towards H1N1 Pandemic. Procedia - Social and Behavioral Sciences 2012;36:77-86 [FREE Full text] [doi: 10.1016/j.sbspro.2012.03.009]
18. Wong C, Tang CS. Practice of habitual and volitional health behaviors to prevent severe acute respiratory syndrome among Chinese adolescents in Hong Kong. J Adolesc Health 2005 Mar;36(3):193-200 [FREE Full text] [doi: 10.1016/j.jadohealth.2004.02.024] [Medline: 15737774]
19. Carpenter C. A meta-analysis of the effectiveness of health belief model variables in predicting behavior. Health Commun 2010 Dec;25(8):661-669 [FREE Full text] [doi: 10.1080/10410236.2010.521906] [Medline: 21153982]
20. Jones C, Jensen JD, Scherr CL, Brown NR, Christy K, Weaver J. The Health Belief Model as an explanatory framework in communication research: exploring parallel, serial, and moderated mediation. Health Commun 2015;30(6):566-576 [FREE Full text] [doi: 10.1080/10410236.2013.873363] [Medline: 25010519]
21. Orji R, Vassileva J, Mandryk R. Towards an effective health interventions design: A extension of the Health Belief Model. Online Journal of Public Health Informatics 4 (3) 2012:A [FREE Full text] [doi: 10.5210/ojphi.v4i3.4321]
22. Ahorsu D, Lin C, Imani V, Saffari M, Griffiths M, Pakpour A. The Fear of COVID-19 Scale: Development and Initial Validation. Int J Ment Health Addiction 2020 Mar 27:1-9 [FREE Full text] [doi: 10.1007/s11469-020-00270-8]
23. BBC News. US students party on spring break despite coronavirus. URL: https://www.bbc.com/news/av/world-us-canada-51955362 [accessed 2020-05-20]
24. Moorman C, Matulich E. A Model of Consumers' Preventive Health Behaviors: The Role of Health Motivation and Health Ability. J Consum Res 1993 Sep;20(2):208-228 [FREE Full text] [doi: 10.1086/209344]

25. Basch C, Hillyer GC, Meleo-Erwin ZC, Jaime C, Mohlman J, Basch CE. Correction: Preventive Behaviors Conveyed on YouTube to Mitigate Transmission of COVID-19: Cross-Sectional Study. JMIR Public Health Surveill 2020 May 06;6(2):e19601 [FREE Full text] [doi: 10.2196/19601] [Medline: 32374718]
26. Bandura A. Social cognitive theory and control over HIV. In: DiClemente R, Peterson J, editors. Preventing AIDS: Theories and methods of behavioral interventions. New York, NY, USA: Plenum Press; 1994:23-59.
27. Strecher V, Champion VL, Rosenstock IM. The Health Belief Model and health behavior. In: Goschman DS, editor. Handbook of Health Behavior Research. New York, NY, USA: Plenum Press; 1997:71-91.
28. Champion VL, Skinner CS. The Health Belief Model. In: Glanz K, Rimer B, Viswanath K, editors. Health Behavior and Health Education. San Francisco, CA, USA: Josey-Bass; 2008:45-65.
29. Lindell M, Whitney D. Accounting for common method variance in cross-sectional research designs. J Appl Psychol 2001 Feb;86(1):114-121 [FREE Full text] [doi: 10.1037/0021-9010.86.1.114] [Medline: 11302223]
30. Jayachandran S, Sharma S, Kaufman P, Raman P. The Role of Relational Information Processes and Technology Use in Customer Relationship Management. Journal of Marketing 2018 Oct 10;69(4):177-192 [FREE Full text] [doi: 10.1509/jmkg.2005.69.4.177]
31. Preacher K, Hayes A. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 2008 Aug;40(3):879-891 [FREE Full text] [doi: 10.3758/brm.40.3.879] [Medline: 18697684]
32. Krieger J, Sarge M. A serial mediation model of message framing on intentions to receive the human papillomavirus (HPV) vaccine: revisiting the role of threat and efficacy perceptions. Health Commun 2013;28(1):5-19 [FREE Full text] [doi: 10.1080/10410236.2012.734914] [Medline: 23330854]
33. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis. New York, NY, USA: Guilford; 2013.
34. Hitlan RT, Zárate MA, Kelly KM, DeSoto MC. Linguistic ostracism causes prejudice: Support for a serial mediation effect. J Soc Psychol 2016;156(4):422-436. [doi: 10.1080/00224545.2015.1119668] [Medline: 26577024]
35. Li AN, Tan HH. What happens when you trust your supervisor? Mediators of individual performance in trust relationships. J Organiz Behav 2012 Jun 25;34(3):407-425. [doi: 10.1002/job.1812]
36. Fenichel EP, Kuminoff NV, Chowell G. Skip the trip: air travelers' behavioral responses to pandemic influenza. PLoS One 2013;8(3):e58249 [FREE Full text] [doi: 10.1371/journal.pone.0058249] [Medline: 23526970]
37. Mao L. Evaluating the combined effectiveness of influenza control strategies and human preventive behavior. PLoS One 2011;6(10):e24706 [FREE Full text] [doi: 10.1371/journal.pone.0024706] [Medline: 22043275]
38. Brug J, Aro AR, Richardus JH. Risk perceptions and behaviour: towards pandemic control of emerging infectious diseases: international research on risk perception in the control of emerging infectious diseases. Int J Behav Med 2009;16(1):3-6 [FREE Full text] [doi: 10.1007/s12529-008-9000-x] [Medline: 19127440]
39. Song J, Song TM, Seo D, Jin D, Kim JS. Social Big Data Analysis of Information Spread and Perceived Infection Risk During the 2015 Middle East Respiratory Syndrome Outbreak in South Korea. Cyberpsychol Behav Soc Netw 2017 Jan;20(1):22-29. [doi: 10.1089/cyber.2016.0126] [Medline: 28051336]
40. Garfin D, Silver R, Holman E. The novel coronavirus (COVID-2019) outbreak: Amplification of public health consequences by media exposure. Health Psychol 2020 May;39(5):355-357 [FREE Full text] [doi: 10.1037/hea0000875] [Medline: 32202824]
41. Wang Y, McKee M, Torbica A, Stuckler D. Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]

Abbreviations

EVT: Expectancy-Value Theory
HBM: Health Belief Model
MERS: Middle East respiratory syndrome
SARS: severe acute respiratory syndrome
SCT: Social Cognitive Theory

Edited by G Fagherazzi; submitted 04.08.20; peer-reviewed by D Chao, W Evans; comments to author 26.10.20; revised version received 21.12.20; accepted 01.02.21; published 17.03.21.

Please cite as:
Roberts JA, David ME
Improving Predictions of COVID-19 Preventive Behavior: Development of a Sequential Mediation Model
J Med Internet Res 2021;23(3):e23218
URL: https://www.jmir.org/2021/3/e23218
doi:10.2196/23218
PMID:33651707

Original Paper

Temporal Dynamics of Public Emotions During the COVID-19 Pandemic at the Epicenter of the Outbreak: Sentiment Analysis of Weibo Posts From Wuhan

Shaobin Yu[1], PhD; David Eisenman[2,3], PhD; Ziqiang Han[1,4], PhD

[1]Department of Public Administration, School of Political Science and Public Administration, Shandong University, Qingdao, China

[2]Division of General Internal Medicine and Health Services Research, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, United States

[3]Center for Public Health and Disasters, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA, United States

[4]Center for Crisis Management Research, Tsinghua University, Beijing, China

Corresponding Author:
Ziqiang Han, PhD
Department of Public Administration
School of Political Science and Public Administration
Shandong University
Binhai Road 72
Jimo
Qingdao, 266237
China
Phone: 86 13210178710
Email: ziqiang.han@sdu.edu.cn

Abstract

Background: The ongoing COVID-19 pandemic has led to an increase in anxiety, depression, posttraumatic stress disorder, and psychological stress experienced by the general public in various degrees worldwide. However, effective, tailored mental health services and interventions cannot be achieved until we understand the patterns of mental health issues emerging after a public health crisis, especially in the context of the rapid transmission of COVID-19. Understanding the public's emotions and needs and their distribution attributes are therefore critical for creating appropriate public policies and eventually responding to the health crisis effectively, efficiently, and equitably.

Objective: This study aims to detect the temporal patterns in emotional fluctuation, significant events during the COVID-19 pandemic that affected emotional changes and variations, and hourly variations of emotions within a single day by analyzing data from the Chinese social media platform Weibo.

Methods: Based on a longitudinal dataset of 816,556 posts published by 27,912 Weibo users in Wuhan, China, from December 31, 2019, to April 31, 2020, we processed general sentiment inclination rating and the type of sentiments of Weibo posts by using pandas and SnowNLP Python libraries. We also grouped the publication times into 5 time groups to measure changes in netizens' sentiments during different periods in a single day.

Results: Overall, negative emotions such as surprise, fear, and anger were the most salient emotions detected on Weibo. These emotions were triggered by certain milestone events such as the confirmation of human-to-human transmission of COVID-19. Emotions varied within a day. Although all emotions were more prevalent in the afternoon and night, fear and anger were more dominant in the morning and afternoon, whereas depression was more salient during the night.

Conclusions: Various milestone events during the COVID-19 pandemic were the primary events that ignited netizens' emotions. In addition, Weibo users' emotions varied within a day. Our findings provide insights into providing better-tailored mental health services and interventions.

(J Med Internet Res 2021;23(3):e27078) doi:10.2196/27078

KEYWORDS

public health emergencies; emotion; infodemiology; temporal dynamics; sentiment analysis; COVID-19

Introduction

Understanding the public's emotional reactions to public health emergencies is essential in at least two ways. First, it is necessary to understand the impact of the public health crisis on the public's mental health and psychological well-being, which are the unneglected dimensions of a human being's health. The ongoing COVID-19 pandemic has led to an increase in anxiety, depression, posttraumatic stress disorder, psychological stress, and stress experienced by the general public in various degrees worldwide [1], similar to the 2003 SARS (severe acute respiratory syndrome) [2], Ebola [3], or Zika outbreaks [4]. Better tailored mental health services and interventions cannot be achieved until we understand the patterns of mental health issues that emerge after a public health crisis; this is especially true in the case of the rapidly spreading COVID-19 [5]. Thus, researchers call for multidisciplinary research for mental health as research priorities [6]. Second, the public's collective emotions during emergencies are important in determining the emergency response to the public health crisis. Understanding the public's emotions and needs and their distribution attributes are critical to develop appropriate public policies, as well as to nudge the public's compliance to these policies and, eventually, respond to the health crises in an effective, efficient, and equitable manner.

There are generally two approaches to capture the public's emotional response to an event. The traditional method involves conducting a large-scale survey, which can be both expensive and time consuming. Moreover, the self-response or retrospective bias is an inevitable criticism of the survey method. Another way is to capture the public's emotional beats on social media platforms, such as Twitter, considering the increasing use of smartphones in recent years. With the rapid development of the internet, smartphones, and social media platforms and apps, an increasing number of individuals use social media as a medium to share their opinions, social activities, and lives. Therefore, social media can be an ideal channel to capture the public's emotions and mental health symptoms [7], especially during the COVID-19 outbreak when physical distancing policies are enforced [8-11].

Since its outbreak, COVID-19 has become an ongoing hot topic on social media. During the early phase of the outbreak, before March 15, 2020, the main topics discussed on Twitter in the English language were the origin of the virus, its sources, transmission, the impact on the economy and society, and the methods of mitigating the risks [12]. Similarly, on the Chinese social media platform Weibo, the primary concerns among users were the origin of the disease, public health control measures, disease symptoms, organizations in charge, health professionals and scientists, education sectors, economic impact, and rumor discussions [13,14]. Since the COVID-19 has lasted for such a long time and is still an ongoing pandemic, the public's concerns and emotional responses have evolved with the pandemic's rapid change, its impacts, and countermeasure policies [15]. The declaration of the COVID-19 outbreak increased anxiety, depression, and indignation emotions among the public, whereas it decreased the public's happiness and life satisfaction [16,17]. The lockdown in Wuhan city in China and Lombardy city in Italy affected the public's expressions on social media; moreover, a higher level of cognitive process was detected among social media users in both cities. Simultaneously, attention to group, religion, and emotions was more prevalent in Wuhan [18]. In India, a comparison between lockdown 2.0 (April 15 to May 3, 2020) and lockdown 3.0 (May 4 to May 17, 2020) revealed that the public's surprise, trust, fear, joy, and anger emotions decreased, but their anticipation, disgust, and sadness emotions increased [19]. However, policies such as physical distancing measures affected different emotions differently. A study from Spain suggested that negative emotions such as anger, fear, and disgust significantly changed, whereas sadness, joy, and uncertainty did not fluctuate much [20]. Meanwhile, positive sentiments, such as trust, anticipation, and joy, increased when discussions concerning reopening emerged on these platforms, although these emotions were mixed with negative sentiments such as fear, sadness, anger, surprise, and disgust [21].

Nevertheless, after reviewing the current sentiment analysis and emotional responses on social media, we found a lack of longitudinal observations from the declaration of the outbreak, strict enforcement of physical distancing and other constraining measures, and the reopening period. Therefore, we employed our ongoing crawled Weibo posts from 27,912 users from Wuhan to capture the emotional dynamics during a specific period of the COVID-19 pandemic, that is, from the beginning of the outbreak to the end of April 2020 (when the whole country reopened). Moreover, we divided the 24 hours of a day into several time periods to explore the emotional variations within a given day, which can be very useful for potential interventions. We used a sentiment classification algorithm that includes 6 types of emotions (ie, like, dislike, surprise, depression, anger, and fear) and analyzed the temporal dynamics of these emotions from 816,556 Weibo posts published by 27,912 users in Wuhan. In this paper, we describe (1) the temporal patterns of emotional fluctuation from the declaration of the COVID-19 outbreak in Wuhan to reopening of the city after the lockdown, (2) significant events during the COVID-19 pandemic that affected users' emotional changes and variations, and (3) hourly variations in the netizens' emotions expressed within a single day.

Methods

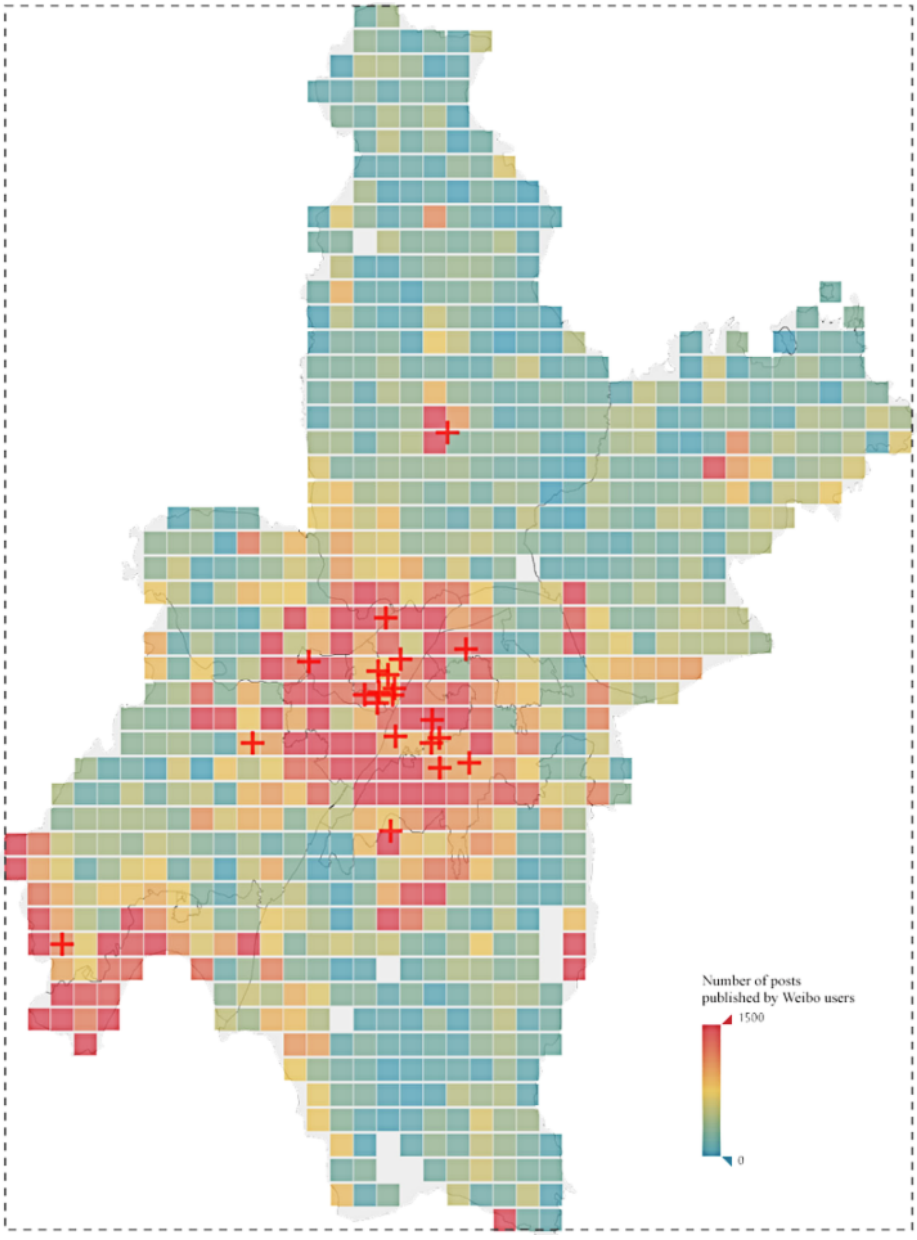
Data Source and Cleaning

Weibo is a social media platform like Twitter. It is the most widely used social networking platform for netizens in China to share and discuss individual opinions, events, and activities publicly. According to the 46th China Statistical Report on Internet Development report released in 2020, there were 940 million active internet users by the end of June 2020, which comprised 67% of all citizens in China [22]. The lockdown and physical distancing measures in response to the COVID-19 pandemic made the use of the internet more prevalent, and the number of Weibo active users reached its record high of 550 million in the first quarter of 2020, according to Weibo's Q1 2020 financial report [23]. Thus, Weibo can be considered as

one of the most appropriate social media channels to detect public emotions toward public events.

We started to collect real-time Weibo posts as soon as a coronavirus outbreak was declared in China on December 31, 2019. For this study, we examined the dataset, including all posts posted by representative Weibo users in Wuhan from December 31, 2019, to April 31, 2020. As shown in Figure 1, we first detected the latitude and longitude of the northernmost (31.36287 N), southernmost (29.97008 N), westernmost (113.70222 E), and easternmost (115.08138 E) locations of Wuhan city. Then, we divided the quadrilateral (dotted line in Figure 1) that covered Wuhan's geographical areas into 50 segments in width and 40 segments in height. Thus, 2000 small grids were generated. Next, 50 Weibo users from each grid were randomly selected from all users located in that grid. The users' geolocations were detected by the geotag used when they posted the message and the location of the place registered per their user profiles. Thus, 100,000 Weibo users were initially selected, from which 88,798 users were retained. The users excluded were either not located in Wuhan or duplicated because they posted multiple times using geotags from different grids or places within Wuhan's geographical area. Lastly, 27,912 users were randomly selected to participate in this study, and all of their 816,556 Weibo posts from December 31, 2019, to April 31, 2020, were collected using a web crawler. For the analysis, we collected (1) Weibo users' profiles; (2) geolocation information of the published Weibo posts; (3) attributes of post likes, date, and time; (4) post contents; and (5) like, reply, and forward counts of the posts.

Figure 1. Heatmap of Weibo users' post frequency by geolocation in Wuhan, China. Red cross symbols on the map indicates Grade-A tertiary hospital in the city.



The original dataset was structured as a CSV (comma-separated values) file and processed using the pandas Python library [24]. In the data cleaning process, non-Unicode characters, empty blank spaces, and symbols were identified by regular expression

patterns and then stripped. Duplicated and invalid posts were also removed. After the data cleaning process, 808,205 valid Weibo posts were selected for further analysis.

Sentiment Analysis

Sentiment analysis is the study of emotions, opinions, appraisals, and attitudes regarding "services, products, individuals, organizations, issues, topics, events and their attributes" [25]. To explore netizens' emotions toward COVID-19, we evaluated both general sentiment inclination and the type of sentiment, as well as their fluctuations during different periods within a day.

The general sentiment inclination rating of Weibo posts was processed by the SnowNLP library for Python [26]. SnowNLP, as one of the most popular Python libraries explicitly developed for Chinese language natural language processing, has a variety of built-in functions such as tokenization, sentiment analysis, text classification, and keyword extraction. SnowNLP also supports a self-trained dataset to improve sentiment rating accuracy for specific types of Weibo posts. Therefore, to measure sentiment ratings, a portion of the dataset was selected as the training data, and then the training result was used to reprocess the whole data and calculate the final sentiment ratings.

Specifically, we randomly selected 50,000 posts from the dataset as the training data. Thereafter, following the instructions of SnowNLP, we manually tagged each post as negative or positive sentiment and then appended the post to corresponding SnowNLP built-in corpus files that were named as neg.txt and pos.txt, respectively. The updated sentiment rating corpus files were then fed into SnowNLP for training. The result was the newly generated sentiment pattern file named sentiment.marshal in the SnowNLP library folder. Based on the training result, all of the 808,205 posts were evaluated, and sentiment ratings were remeasured. The result was a score ranging from 1 to 10 for each post, denoting the most negative to the most positive sentiments. We then divided each sentiment rating value by 10, and the final sentiment ratings ranged from 0.1 to 1.

Sentiment Type Analysis

To properly align each post's sentiment type, this study used the DUTIR Emotion Ontology set, developed and maintained by Prof Lin Hongfei and his team at Dalian University of Technology Institute of Information Retrieval (DUTIR), as the emotion lexicon resource [27]. Their study compared the tokenized words with the DUTIR Emotion Ontology set and generated a list of words with their corresponding emotional categories for each Weibo post. After that, the word in the emotional category with the most frequent occurrences was treated as the post's baseline sentiment type. Then, each Weibo post was tagged with a sentiment type—depression, like, angry, dislike, surprise, or fear.

Both clinical and self-reported studies indicate that people's moods, such as depression [28,29] or rumination [30], vary within a day. To measure netizens' sentiment changes during different periods in a day, we grouped the hours into 5 time groups: early morning (04:00 to 08:00), morning (08:00 to 12:00), afternoon (12:00 to 17:00), evening (17:00 to 20:00), night (20:00 to 00:00), and late night (00:00 to 04:00). We grouped all of the general sentiment inclinations by day and time using the pandas library's *groupby* method and counted each group's frequencies. Based on the result, we calculated the ratio of each sentiment type in the specific period, that is, the percentage of occurrences of each sentiment type in the specific time group.

Results

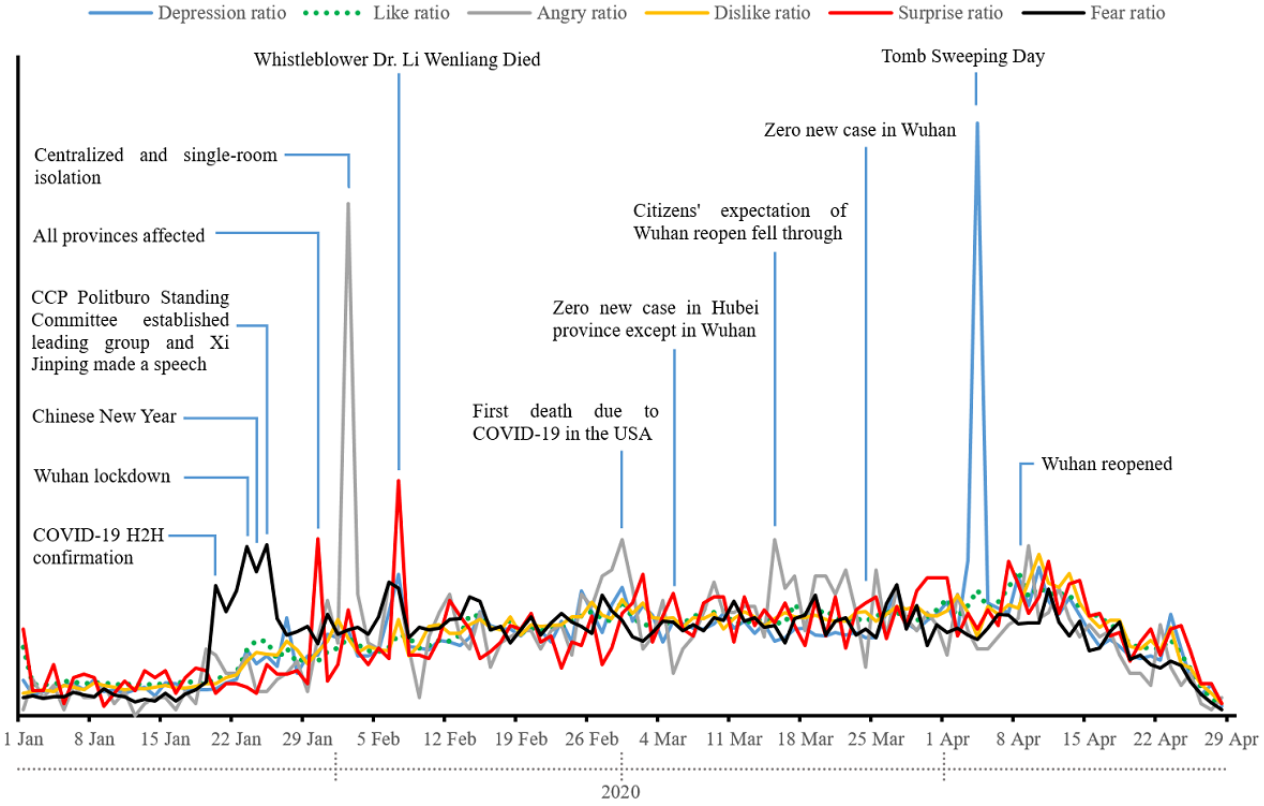
Geolocation Distribution of Weibo Users and Post Frequencies

First, we plotted all of the Weibo users on the map of Wuhan according to their geolocation information. We drew a heatmap overlaid on the city map representing the total number of Weibo posts by users during the whole period. The final output shows the frequency of posts in different locations of Wuhan. Additionally, we plotted the representative Grade-A tertiary hospitals in Wuhan on the map to better understand the relationship between Weibo users' post patterns and their proximity to neighboring hospitals. We used locations of the major hospitals as a proxy of the places where newly confirmed COVID-19 cases occurred, and the distances can reflect the proximity to these places and citizens' posting behaviors. The map demonstrates a clear correlation pattern between the number of Weibo posts published by users and their physical distance to major hospitals.

Temporal Evolution of Sentiment Types

The temporal evolution of different sentiment types is illustrated in Figure 2. The higher the value, the stronger the type of sentiment. Overall, fear, anger, and surprise were the most dominant emotions, whereas like, dislike and depression were the less salient emotions. Similar variations and dynamics were observed for the negative emotions during the analysis. The emotional responses can be broadly differentiated into three periods. First, there was a sharp increase in all sentiments after the confirmation of human-to-human transmission of COVID-19 on January 20, 2020; thereafter, this increasing trend remain steady until February 7, 2020, when the young whistleblower Dr Li Wenliang passed away. Second, the number and variations of posts on Weibo were considerably higher shortly after the confirmation of human-to-human transmission of the virus compared with posts published later. Moreover, a subtle and smooth pattern was maintained across all emotions, with fewer variations noted until Wuhan reopened after April 8, 2020. The third period was after the lockdown was lifted (ie, reopening); a significant decrease of posts and all emotions was observed during this period.

Figure 2. Line chart of the temporal evolution of sentiment types based on Weibo posts from Wuhan, China.

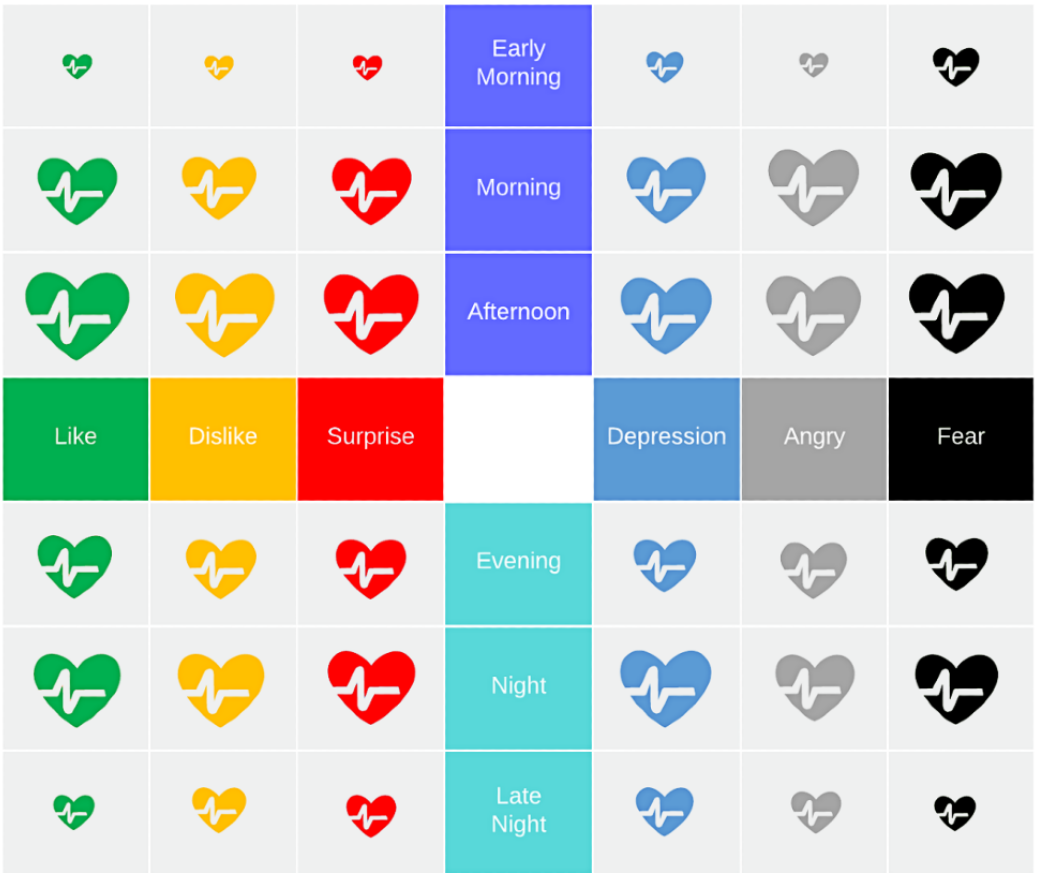


The result also demonstrated that Wuhan residents' emotions were significantly affected by milestone events. Fear was the dominant emotion in the first week; it reached the first peak on January 20, 2020, when human-to-human transmission of COVID-19 was confirmed and then increased to a higher peak on January 23, 2020, with the declaration of the city entering a lockdown. Another peak for fear was recorded one day later when President Xi deployed the emergency response on the first day of the Chinese New Year. With regard to the other emotions, the first peak for depression was reached shortly after the first day of the Chinese New Year. The first peak of surprise was recorded on January 30, 2020, when Tibet, the last province of China to report a COVID-19 outbreak, reported new COVID-19 cases, indicating that all the provinces of Mainland China had reported new positive cases of COVID-19 that day. The first and the highest peak for anger appeared on February 2, 2020, when all infected or suspected patients, those with symptoms similar to COVID-19, or close contacts of infected patients were required to be collectively isolated. The emotions declined on February 5, 2020, when President Xi gave another concrete national response direction. Thereafter, all negative emotions (ie, fear, anger, surprise, and depression) increased and reached a new peak on February 7, 2020, when the public recognized the young doctor Li Wenliang passed away. After these events, all the emotions maintained a steady trend at a low level except for two peaks observed for anger. One peak was observed on February 29, 2020, when the first COVID-19–related death occurred in the United States, and the local officials of Wuhan declared that the city had adequate supplies for food and living goods that can last at least 1 month. The other peak was observed on March 15, 2020, when the public's expectation of the city reopening did not come true. The news about zero new confirmed COVID-19 cases, either in Wuhan or in other places within Hubei province except for Wuhan, had little influence on the residents' emotions. The depression emotion had reached the highest peak and was the dominant emotion around the Tomb Sweeping Day (April 4, 2020), the traditional memorial festival in the Chinese culture for the reverence of ancestors and the passed ones.

Intraday Distribution of Sentiment Types

The variations of emotions within a day are illustrated in Figure 3. We computed the mean intensity of each sentiment type during different periods (ie, early morning, morning, afternoon, evening, night, and late night) and presented them in a matrix diagram. Afternoon (12:00-17:00) and night (20:00-00:00) were the primary periods when people use social media to express their feelings during the COVID-19 epidemic, followed by morning (08:00-12:00), evening (17:00-20:00), late night (00:00-04:00), and early morning (04:00-08:00). In the morning period, fear and anger were the two most dominant emotions. In the afternoon period, like and dislike emotions were the highest (or most dominant), although most emotions expressed on social media were noted during this period. During the night period, depression was the most dominant emotion.

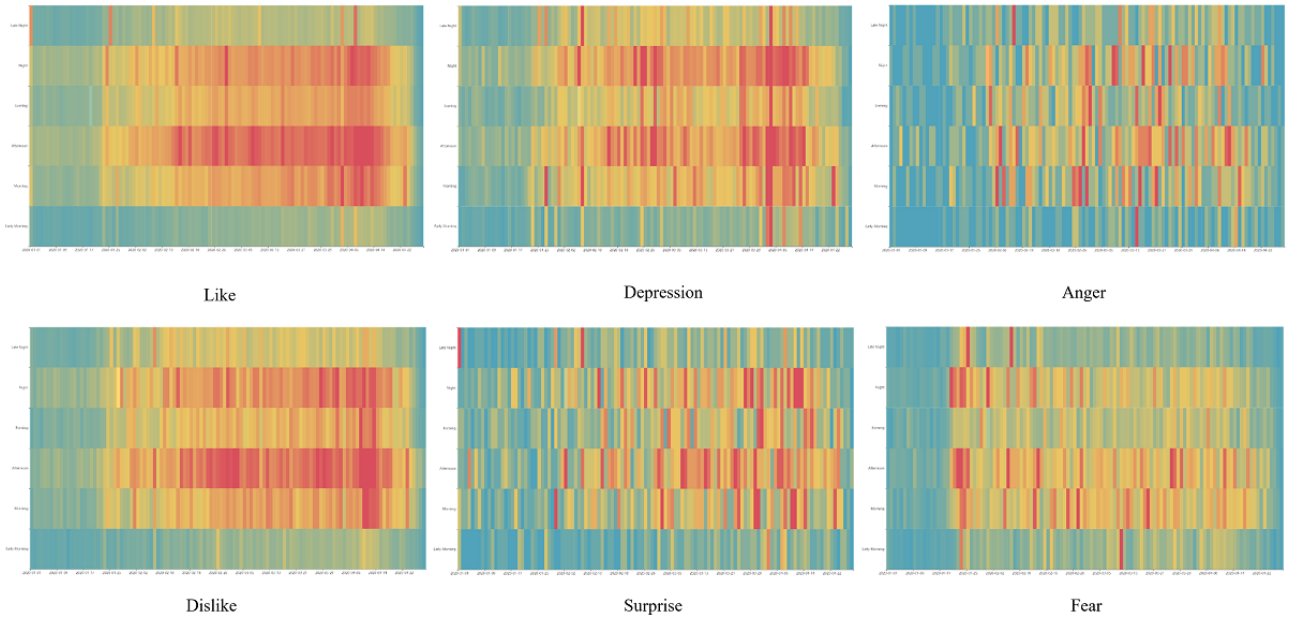
Figure 3. Distribution of different types of sentiments across different time periods used for analysis. The bigger the heart symbol, the stronger the corresponding emotion was during the given time period.



Temporal Evolution of Sentiment Types During Different Time Periods

To better understand the temporal distribution of emotions, we used heatmaps to demonstrate the distribution of each emotion by days and hours within a given day (Figure 4). A stronger emotion is presented in a darker red color. The relatively gentle emotions such as like and dislike were consistent during the lockdown period, especially in the afternoon and night periods. The relatively stronger emotions such as anger and fear appeared periodically, likely triggered by different major events. The emotion depression was also prevalent over time, especially after the young whistleblower doctor passed away and on the Tomb Sweeping day, commemorated in honor of those who have passed away.

Figure 4. Temporal changes in sentiment type by days (horizontal) and hours (vertical).



Discussion

Principal Findings

In this study, we followed the emotional changes of 27,912 social media (Weibo) users from Wuhan, China, from December 31, 2019, to April 31, 2020, covering all the stages before the COVID-19 outbreak, declaration of the outbreak, the lockdown, and the subsequent reopening of the city. Specifically, we analyzed the temporary dynamics of users' sentiments, including like, dislike, surprise, depression, anger, and fear, expressed through the Weibo platform and displayed these emotions across daily and hourly periods. There are at least three noteworthy findings from this study.

First, negative emotions such as surprise, fear, and anger were the dominant emotions on social media since the confirmation of the human-to-human transmission of the virus. In the very early stage of the outbreak (ie, first 2 weeks), fear was the dominant emotional response, when the virus was first detected, and human knowledge about this virus and its impact was minimal. Anger is another dominant emotion that is easily influenced by milestone events. An earlier study has demonstrated that anger was more influential on Weibo than emotions like joy or sadness [31], and a similar trend was observed during the COVID-19 outbreak. The other emotions—like, dislike, and depression—mildly fluctuated in the first half-year of 2020, and all emotions decreased significantly when the city was reopened, but there was a short-term increase in these emotions several days immediately after the reopening of the city post the lockdown. Our finding is similar to a previous study that used a search index of psychological keywords from residents in Wuhan; the search of keywords specific to fear and psychological counseling increased significantly in the first month after the lockdown, compared with the same period in 2019, whereas this was not true for keywords specific to depression and insomnia [32].

Another novel finding of our study is that we detected hourly variations in the emotions within a day, which researchers have rarely paid attention to before. We found that netizens from Wuhan expressed more emotions in the afternoons (12:00-17:00) and nights (20:00-00:00). The emotions anger and fear were more prevalent in the mornings and afternoons, whereas the emotion depression was relatively more prevalent during nighttime. This observation is similar to those of prior studies that have investigated the diurnal variations of mental health–related behaviors. Google search data from Finland shows that depression-related query volumes started to increase in the late evenings and reached a peak around midnight [33], although Canadian police records indicate that the call volumes for mental health service generally peaked in the mid-afternoon [34]. Our study using social media data reconfirms the hourly variations of mental health symptoms, and this finding has potential contribution to the implementation of mental health intervention strategies in the future.

Moreover, the pandemic's evolution and the response measures significantly affected people's emotions, especially anger, in the whole process. It was observed that the general quarantine measures did not have a negative psychological effect on the affected individuals in case of the less-deadly disease like H1N1 [35], but this was not true for the more-deadly disease such as SARS [36]. However, COVID-19—as a collective stressor—reduced the general public's mental health and psychological well-being worldwide [37-40]. Different concerns during different stages of the pandemic evolution are associated with different emotions on social media. For example, fear could be related to the shortage of medical and test supplies; anger could be associated with xenophobia, at first, and with the stay-at-home orders, later; sadness could cooccur with the topics of losing friends or family members; and joy could be associated with words of gratitude and good health [15]. For residents of Wuhan, fear primarily occurred during the first 2 weeks of the outbreak, and the confirmation of human-to-human transmission of the disease, the city's lockdown, and the declaration of the national response by President Xi; moreover, the passing away of the whistleblower Dr Li caused the fear emotion to peak. With the occurrence of positive COVID-19 cases in all provinces of China, Dr Li's death led to a peak of the surprise emotion. The anger emotion first reached an outstandingly high level when a decision was made to quarantine and isolate all infected patients and their close contacts in Wuhan. Other peaks during the process were observed on the day the first confirmed death due to COVID-19 was reported in the USA, which was also the day when the Wuhan city government declared that the city had a full capacity of logistics and supplies, and the day when the public's expectation of city reopening failed. Moreover, the depression emotion exploded on Memorial Day and reached another peak when Dr Li passed away. These findings suggest that public health emergency response strategies should be tailored to the dynamics and evolution of the disease and the public's response [41].

Furthermore, our sampling and data collection can provide methodology insights for studies using social media data. Unlike the studies using keyword search methods, those based on the social media platform's feedback through application programming interfaces [42], those exploring only a small number of social media users, or those covering only a short period [16,18,43], we used a relatively sophisticated sampling method and developed a "longitudinal" dataset of social media posts published by a large sample of respondents (Weibo users). Our analysis and visualization methods can also prove to be a valuable contribution to future studies investigating and presenting netizens' opinions and behavioral changes.

Limitations

Nevertheless, there are at least two limitations of our study. The first is that we did not link Weibo users' profiles with their posts and emotional dynamics. If we can detect the demographic variations of these emotional changes, it may be better to detect the psychological needs of the residents during a public health emergency response, especially in the context of the lockdown of a city. However, there could be ethical and privacy concerns of including the detailed profile information of social media users [44,45], particularly considering those who are unwilling to reveal their mental health status. Hence, we anonymized all users' information in our analysis. Second, we did not include the users' behavioral changes, including both protective behaviors and those that have a negative impact (eg, alcohol

consumption) in this analysis. We also did not cover posttraumatic stress disorder and anxiety emotions, which can also be prevalent and vital during the COVID-19 pandemic [46,47].

Conclusions

In this study, we investigated temporal emotions (like, dislike, surprise, fear, depression, and anger) and their dynamics from nearly 30,000 Weibo users in Wuhan from December 31, 2019, to April 31, 2020, covering different periods, ranging from before the COVID-19 outbreak, the resulting city-wide lockdown, and the subsequent reopening of the city. Milestone events were the primary triggers that ignited various emotions during this period. Negative emotions such as surprise, fear, and anger were the most salient emotions on the Chinese social media platform Weibo. The emotions also varied within a given day. All emotions analyzed were more prevalent in the afternoon and night, but fear and anger were more likely to appear in the morning and afternoon, and depression was more salient during the night. Our finding can potentially contribute to the implementation of mental health intervention strategies in the future. In addition, our analysis and visualization methods can also prove to be a valuable contribution to future studies investigating and concerning netizens' opinions and behavioral changes.

Acknowledgments

This research was supported by the National Social Science Foundation of China (20ZD&160) and the National Natural Science Foundation of China (72042008). The funding sources had no involvement in the study design, data collection, analysis, interpretation of data, writing of the report, or decision to submit the manuscript for publication.

Authors' Contributions

ZH contributed to the overall research design and manuscript preparation. SY contributed to the data collection, methodology, and data analyses. All authors contributed to reviewing, editing, revision, and submission of the manuscript for publication.

Conflicts of Interest

None declared.

References

1. Xiong J, Lipsitz O, Nasri F, Lui LMW, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: a systematic review. J Affect Disord 2020 Dec 01;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
2. Reynolds D, Garay J, Deamond S, Moran M, Gold W, Styra R. Understanding, compliance and psychological impact of the SARS quarantine experience. Epidemiol Infect 2008 Jul;136(7):997-1007. [doi: 10.1017/S0950268807009156] [Medline: 17662167]
3. Cénat JM, Mukunzi JN, Noorishad P, Rousseau C, Derivois D, Bukaka J. A systematic review of mental health programs among populations affected by the Ebola virus disease. J Psychosom Res 2020 Feb 13;131:109966. [doi: 10.1016/j.jpsychores.2020.109966] [Medline: 32087433]
4. Mamidi R, Miller M, Banerjee T, Romine W, Sheth A. Identifying key topics bearing negative sentiment on Twitter: insights concerning the 2015-2016 Zika epidemic. JMIR Public Health Surveill 2019 Jun 04;5(2):e11036 [FREE Full text] [doi: 10.2196/11036] [Medline: 31165711]
5. Li W, Yang Y, Liu Z, Zhao Y, Zhang Q, Zhang L, et al. Progression of mental health services during the COVID-19 outbreak in China. Int J Biol Sci 2020;16(10):1732-1738 [FREE Full text] [doi: 10.7150/ijbs.45120] [Medline: 32226291]
6. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
7. Gupta A, Katarya R. Social media based surveillance systems for healthcare using machine learning: a systematic review. J Biomed Inform 2020 Aug;108:103500 [FREE Full text] [doi: 10.1016/j.jbi.2020.103500] [Medline: 32622833]
8. Guntuku SC, Sherman G, Stokes DC, Agarwal AK, Seltzer E, Merchant RM, et al. Tracking mental health and symptom mentions on Twitter during COVID-19. J Gen Intern Med 2020 Sep;35(9):2798-2800 [FREE Full text] [doi: 10.1007/s11606-020-05988-8] [Medline: 32638321]
9. Jia F, Liu X. Using social media to analyze public psychological status during the recovery period of COVID-19 in China. J Public Health (Oxf) 2020 Sep 03:fdaa152 [FREE Full text] [doi: 10.1093/pubmed/fdaa152] [Medline: 32880617]
10. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter discussions and emotions about the COVID-19 pandemic: machine learning approach. J Med Internet Res 2020 Nov 25;22(11):e20550 [FREE Full text] [doi: 10.2196/20550] [Medline: 33119535]
11. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study. JMIR Public Health Surveill 2020 Nov 11;6(4):e21978 [FREE Full text] [doi: 10.2196/21978] [Medline: 33108310]

12. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of Tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
13. Wang T, Lu K, Chow KP, Zhu Q. COVID-19 sensing: negative sentiment analysis on social media in China via BERT model. IEEE Access 2020;8:138162-138169. [doi: 10.1109/access.2020.3012595]
14. Zhu B, Zheng X, Liu H, Li J, Wang P. Analysis of spatiotemporal characteristics of big data on social media sentiment with COVID-19 epidemic topics. Chaos Solitons Fractals 2020 Nov;140:110123 [FREE Full text] [doi: 10.1016/j.chaos.2020.110123] [Medline: 32834635]
15. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: analysis of Twitter trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
16. Li S, Wang Y, Xue J, Zhao N, Zhu T. The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users. Int J Environ Res Public Health 2020 Mar 19;17(6):2032 [FREE Full text] [doi: 10.3390/ijerph17062032] [Medline: 32204411]
17. Meléndez JC, Satorres E, Reyes-Olmedo M, Delhom I, Real E, Lora Y. Emotion recognition changes in a confinement situation due to COVID-19. J Environ Psychol 2020 Dec;72:101518. [doi: 10.1016/j.jenvp.2020.101518]
18. Su Y, Xue J, Liu X, Wu P, Chen J, Chen C, et al. Examining the impact of COVID-19 lockdown in Wuhan and Lombardy: A psycholinguistic analysis on Weibo and Twitter. Int J Environ Res Public Health 2020 Jun 24;17(12):4552 [FREE Full text] [doi: 10.3390/ijerph17124552] [Medline: 32599811]
19. Chehal D, Gupta P, Gulati P. COVID-19 pandemic lockdown: An emotional health perspective of Indians on Twitter. Int J Soc Psychiatry 2020 Jul 07:20764020940741 [FREE Full text] [doi: 10.1177/0020764020940741] [Medline: 32633185]
20. Iglesias-Sánchez PP, Vaccaro Witt GF, Cabrera FE, Jambrino-Maldonado C. The contagion of sentiments during the COVID-19 pandemic crisis: The case of isolation in Spain. Int J Environ Res Public Health 2020 Aug 14;17(16):5918 [FREE Full text] [doi: 10.3390/ijerph17165918] [Medline: 32824110]
21. Samuel J, Rahman MM, Ali GGMN, Samuel Y, Pelaez A, Chong PHJ, et al. Feeling positive about reopening? New normal scenarios from COVID-19 US reopen sentiment analytics. IEEE Access 2020;8:142173-142190. [doi: 10.1109/access.2020.3013933]
22. The 46th China Statistical Report on Internet Development in China. China Internet Network Information Center. Beijing; 2020 Sep. URL: https://cnnic.com.cn/IDR/ReportDownloads/202012/P020201201530023411644.pdf [accessed 2021-03-10]
23. Weibo Reports First Quarter 2020 Unaudited Financial Results. 2020 May 19. URL: http://ir.weibo.com/news-releases/news-release-details/weibo-reports-first-quarter-2020-unaudited-financial-results [accessed 2021-01-18]
24. pandas - Python Data Analysis Library. URL: https://pandas.pydata.org/ [accessed 2020-12-14]
25. D'Andrea A, Ferri F, Grifoni P, Guzzo T. Approaches, tools and applications for sentiment analysis implementation. IJCA 2015 Sep 17;125(3):26-33. [doi: 10.5120/ijca2015905866]
26. Wang R. isnowfy/snownlp - Python library for processing Chinese text. GitHub. URL: https://github.com/isnowfy/snownlp [accessed 2021-02-14]
27. Xu L. Constructing the affective lexicon ontology. J China Soc Entific Tech Inf 2008;27(2):180-185 [FREE Full text]
28. Wirz-Justice A. Diurnal variation of depressive symptoms. Dialogues Clin Neurosci 2008;10(3):337-343 [FREE Full text] [Medline: 18979947]
29. Murray G. Diurnal mood variation in depression: a signal of disturbed circadian function? J Affect Disord 2007 Sep;102(1-3):47-53. [doi: 10.1016/j.jad.2006.12.001] [Medline: 17239958]
30. Takano K, Tanno Y. Diurnal variation in rumination. Emotion 2011 Oct;11(5):1046-1058. [doi: 10.1037/a0022757] [Medline: 21517158]
31. Fan R, Zhao J, Chen Y, Xu K. Anger is more influential than joy: sentiment correlation in weibo. PLoS One 2014;9(10):e110184 [FREE Full text] [doi: 10.1371/journal.pone.0110184] [Medline: 25333778]
32. Chen H, Zhang K. Insight into the psychological problems on the epidemic of COVID-19 in China by online searching behaviors. J Affect Disord 2020 Nov 01;276:1093-1094 [FREE Full text] [doi: 10.1016/j.jad.2020.07.128] [Medline: 32771861]
33. Tana JC, Kettunen J, Eirola E, Paakkonen H. Diurnal Variations of Depression-Related Health Information Seeking: Case Study in Finland Using Google Trends Data. JMIR Ment Health 2018 May 23;5(2):e43 [FREE Full text] [doi: 10.2196/mental.9152] [Medline: 29792291]
34. Vaughan A, Wuschke K, Hewitt A, Hodgkinson T, Andresen M, Brantingham P, et al. Variations in Mental Health Act calls to police: an analysis of hourly and intra-week patterns. PIJPSM 2018 Feb 12;41(1):58-69. [doi: 10.1108/pijpsm-10-2016-0153]
35. Wang Y, Xu B, Zhao G, Cao R, He X, Fu S. Is quarantine related to immediate negative psychological consequences during the 2009 H1N1 epidemic? Gen Hosp Psychiatry 2011;33(1):75-77. [doi: 10.1016/j.genhosppsych.2010.11.001] [Medline: 21353131]

36. Hawryluck L, Gold WL, Robinson S, Pogorski S, Galea S, Styra R. SARS control and psychological effects of quarantine, Toronto, Canada. Emerg Infect Dis 2004 Jul;10(7):1206-1212 [FREE Full text] [doi: 10.3201/eid1007.030703] [Medline: 15324539]
37. Holingue C, Badillo-Goicoechea E, Riehm KE, Veldhuis CB, Thrul J, Johnson RM, et al. Mental distress during the COVID-19 pandemic among US adults without a pre-existing mental health condition: Findings from American trend panel survey. Prev Med 2020 Oct;139:106231 [FREE Full text] [doi: 10.1016/j.ypmed.2020.106231] [Medline: 32758507]
38. Holman EA, Thompson RR, Garfin DR, Silver RC. The unfolding COVID-19 pandemic: a probability-based, nationally representative study of mental health in the United States. Sci Adv 2020 Oct 6;6(42):eabd5390 [FREE Full text] [doi: 10.1126/sciadv.abd5390] [Medline: 32948511]
39. Pierce M, Hope H, Ford T, Hatch S, Hotopf M, John A, et al. Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population. Lancet Psychiatry 2020 Oct;7(10):883-892 [FREE Full text] [doi: 10.1016/S2215-0366(20)30308-4] [Medline: 32707037]
40. Ran L, Wang W, Ai M, Kong Y, Chen J, Kuang L. Psychological resilience, depression, anxiety, and somatization symptoms in response to COVID-19: A study of the general population in China at the peak of its epidemic. Soc Sci Med 2020 Oct;262:113261 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113261] [Medline: 32758794]
41. Han Z, Ai X, Veuthey J. How is COVID-19 being contained in China? An observational study on the local level. Disaster Med Public Health Prep 2020 Dec 22;8:1-26 (forthcoming) [FREE Full text] [doi: 10.1017/dmp.2020.491]
42. Zhao Y, Cheng S, Yu X, Xu H. Chinese public's attention to the COVID-19 epidemic on social media: Observational descriptive study. J Med Internet Res 2020 May 04;22(5):e18825 [FREE Full text] [doi: 10.2196/18825] [Medline: 32314976]
43. Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: Using latent Dirichlet allocation for topic modeling on Twitter. PLoS One 2020;15(9):e0239441 [FREE Full text] [doi: 10.1371/journal.pone.0239441] [Medline: 32976519]
44. Hunter RF, Gough A, O'Kane N, McKeown G, Fitzpatrick A, Walker T, et al. Ethical issues in social media research for public health. Am J Public Health 2018 Mar;108(3):343-348. [doi: 10.2105/AJPH.2017.304249] [Medline: 29346005]
45. McKee R. Ethical issues in using social media for health and health care research. Health Policy 2013 May;110(2-3):298-301. [doi: 10.1016/j.healthpol.2013.02.006] [Medline: 23477806]
46. Daughton AR, Paul MJ. Identifying protective health behaviors on Twitter: Observational study of travel advisories and Zika virus. J Med Internet Res 2019 May 13;21(5):e13090 [FREE Full text] [doi: 10.2196/13090] [Medline: 31094347]
47. Pedrosa AL, Bitencourt L, Fróes ACF, Cazumbá MLB, Campos RGB, de Brito SBCS, et al. Emotional, behavioral, and psychological impact of the COVID-19 pandemic. Front Psychol 2020 Oct;11:566212 [FREE Full text] [doi: 10.3389/fpsyg.2020.566212] [Medline: 33117234]

Abbreviations

DUTIR: Dalian University of Technology, Institute of Information Retrieval
SARS: severe acute respiratory syndrome

Edited by C Basch; submitted 10.01.21; peer-reviewed by A Alasmari, R Subramaniyam; comments to author 10.02.21; revised version received 17.02.21; accepted 01.03.21; published 18.03.21.

Please cite as:
Yu S, Eisenman D, Han Z
Temporal Dynamics of Public Emotions During the COVID-19 Pandemic at the Epicenter of the Outbreak: Sentiment Analysis of Weibo Posts From Wuhan
J Med Internet Res 2021;23(3):e27078
URL: https://www.jmir.org/2021/3/e27078
doi:10.2196/27078
PMID:33661755

Original Paper

Revealing Opinions for COVID-19 Questions Using a Context Retriever, Opinion Aggregator, and Question-Answering Model: Model Development Study

Zhao-Hua Lu[1], PhD; Jade Xiaoqing Wang[1], PhD; Xintong Li[2], PhD

[1]Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, United States

[2]Department of Linguistics, The Ohio State University, Columbus, OH, United States

Corresponding Author:
Zhao-Hua Lu, PhD
Department of Biostatistics
St. Jude Children's Research Hospital
MS 768, Room R6006
262 Danny Thomas Place
Memphis, TN, 38105-3678
United States
Phone: 1 901 595 2714
Email: zhaohua.lu@stjude.org

Abstract

Background: COVID-19 has challenged global public health because it is highly contagious and can be lethal. Numerous ongoing and recently published studies about the disease have emerged. However, the research regarding COVID-19 is largely ongoing and inconclusive.

Objective: A potential way to accelerate COVID-19 research is to use existing information gleaned from research into other viruses that belong to the coronavirus family. Our objective is to develop a natural language processing method for answering factoid questions related to COVID-19 using published articles as knowledge sources.

Methods: Given a question, first, a BM25-based context retriever model is implemented to select the most relevant passages from previously published articles. Second, for each selected context passage, an answer is obtained using a pretrained bidirectional encoder representations from transformers (BERT) question-answering model. Third, an opinion aggregator, which is a combination of a biterm topic model and k-means clustering, is applied to the task of aggregating all answers into several opinions.

Results: We applied the proposed pipeline to extract answers, opinions, and the most frequent words related to six questions from the COVID-19 Open Research Dataset Challenge. By showing the longitudinal distributions of the opinions, we uncovered the trends of opinions and popular words in the articles published in the five time periods assessed: before 1990, 1990-1999, 2000-2009, 2010-2018, and since 2019. The changes in opinions and popular words agree with several distinct characteristics and challenges of COVID-19, including a higher risk for senior people and people with pre-existing medical conditions; high contagion and rapid transmission; and a more urgent need for screening and testing. The opinions and popular words also provide additional insights for the COVID-19–related questions.

Conclusions: Compared with other methods of literature retrieval and answer generation, opinion aggregation using our method leads to more interpretable, robust, and comprehensive question-specific literature reviews. The results demonstrate the usefulness of the proposed method in answering COVID-19–related questions with main opinions and capturing the trends of research about COVID-19 and other relevant strains of coronavirus in recent years.

(J Med Internet Res 2021;23(3):e22860) doi:10.2196/22860

KEYWORDS

natural language processing; question-answering systems; language summarization; machine learning; life and medical sciences; COVID-19; public health; coronavirus literature

Introduction

COVID-19 is an infectious disease that emerged in late 2019 and has resulted in an ongoing pandemic [1]. COVID-19 is caused by SARS-CoV-2, which belongs to a family of viruses known as coronaviruses. Coronaviruses are enveloped positive-sense single-stranded ribonucleic acid viruses. Recently discovered coronaviruses include severe acute respiratory syndrome coronavirus (SARS-CoV) [2] and Middle East respiratory syndrome coronavirus (MERS-CoV) [3]. COVID-19 has challenged global public health because it is highly contagious and can be lethal. Consequently, numerous ongoing and recently published studies have emerged in the recent literature. However, research regarding COVID-19 is largely ongoing and inconclusive.

A potential approach that could accelerate COVID-19 research is to use information from the existing research into other viruses that belong to the coronavirus family. To address the COVID-19 pandemic challenges, the US government and several leading research groups have created the COVID-19 Open Research Dataset (CORD-19) [4], which includes scholarly articles about COVID-19 and related coronaviruses. However, it would be difficult for the medical research community to keep up with the continuously growing literature. The application of natural language processing (NLP), together with relevant statistical techniques, can help to generate new insights and support the ongoing fight against COVID-19.

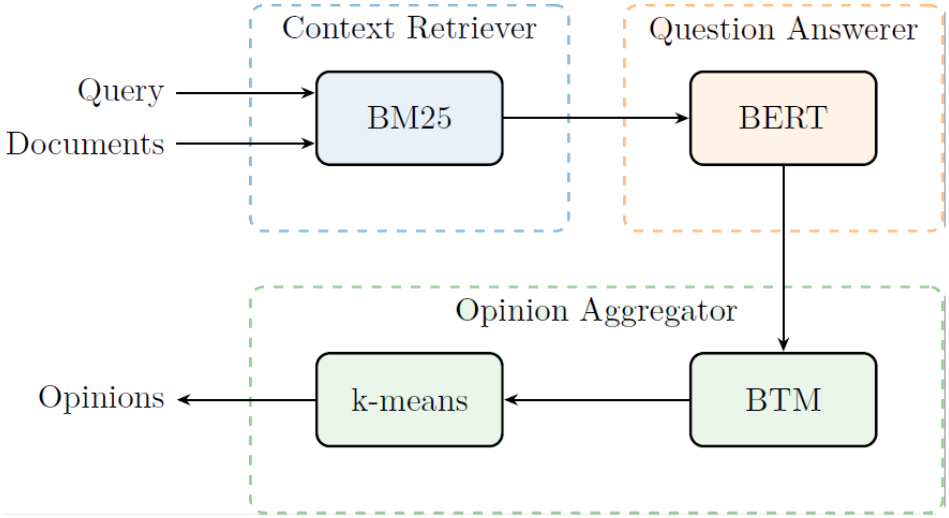
Our motivation is to address the two important questions raised by the CORD-19 Challenge: (1) use natural language processing to find answers to questions within, and connect insights across, the CORD-19 database in support of the ongoing COVID-19 response efforts worldwide; and (2) help the medical community keep up with the rapid increase in COVID-19 literature. Our goal is to make use of these techniques to distill information from the CORD-19 data set and generate answers and summaries for a variety of questions from different perspectives. In addition, we will demonstrate the longitudinal trends of the research related to the coronavirus family and compare the answers with our existing understanding of COVID-19 and other viruses in the coronavirus family.

Methods

Overview

We propose a pipeline to automatically answer various COVID-19–related questions from the literature and aggregate the answers into several main opinions. The pipeline consists of three components as shown in Figure 1: (1) a context retriever module for finding relevant passages from the articles, (2) a question-answering model for extracting answers from a single context passage, and (3) an opinion aggregator for clustering answers into opinions. The popularity of an opinion is quantified by the number of supportive answers in the literature. We show the trends of the related research by investigating the research trends quantified by the longitudinal distributions of the opinions based on the answers to each question.

Figure 1. Illustration of the pipeline. BERT: bidirectional encoder representations from transformers; BTM: biterm topic model.



Related Work

The release of the CORD-19 data set and the call for using NLP techniques to exploit CORD-19 have inspired considerable research effort toward building systems that can help researchers to explore valuable information related to COVID-19. Some earlier contributions, including CORD-19 Search [5], tmCOVID [6], WellAI COVID-19 Research Tool [7], and Covidex [8], focus on search engine development or identification of medical concepts. On the other hand, with the emerging requests from both the medical research community and society in general to find answers to various questions regarding COVID-19, systems that can provide reliable answers to COVID-19–related questions using the latest COVID-19 research are urgently needed.

Some very recent developments in answering systems built on top of the CORD-19 corpus include CovidQA [9], CAiRE-COVID [10], and CO-Search [11]. These methods tried to generate answers for COVID-19–related questions by locating a fraction of the documents in CORD-19 or summarizing the documents into a final answer through text generation (eg, BART [12]).

Compared to these very recent works, our research has several unique contributions. First, for a user query, our proposed system can automatically aggregate the answers from the most related articles into a few main opinions, which helps researchers to acquire clear directions. Second, the opinions are generated with popularity measurement as the number of supportive answers in the selected articles. Given that COVID-19 is being actively researched, multiple opinions accompanied by popularity measurements may answer open questions better than the definitive answers produced by the other methods. Third, our research pays more attention to data analyses and interpretations, while other research focuses on model performance evaluation. Nonetheless, we provide clear justifications for the selected components in our pipeline. Below, we describe our pipeline's three components: a context retriever, a question-answering model, and an opinion aggregator (Figure 1).

Context Retriever

Using a neural model to extract answers from passages of the entire literature is costly, and it is more efficient to filter the irrelevant passages out first. Chen et al [13] used term frequency–inverse document frequency (TF-IDF) first to narrow the search space so that the following steps could focus only on the relevant passages. Likewise, we use BM25 [14,15] as an enhanced version of TF-IDF to quantify the relevance of the passages deterministically and select the relevant passages. Compared to the linear term frequency score in TF-IDF, BM25 alleviates the imbalance impact of high term frequency by limiting the range of the term frequency score to represent term saturation (ie, after the term appears a sufficient number of times, a higher frequency of the term does not affect the relevance of a passage with the term). Furthermore, BM25 uses the specific document length relative to the average document length to determine the sufficient term number for saturation, resulting in a more flexible and stable context retriever. Specifically, the question and the passages are regarded as a bag of words, and they are compared through BM25 similarity. For each question, all passages are ranked according to their similarity values with the question, and only the top passages are selected.

Question-Answering Model

Given a question and a passage, we apply the bidirectional encoder representations from transformers (BERT) [16,17] question-answering model pretrained by Huggingface [18] to each passage selected by the context retriever to predict the location and span of answer text in each passage. We chose the large, uncased version of BERT. Specifically, the question and the passage are packed into a single sequence separated by a segment embedding. Their contextual representations are computed by BERT. The start position of the answer in the passage is the word with the contextural representation that leads to the largest inner product with the start vector in the pretrained BERT question-answering model. The end position is similarly calculated using the end vector in the pretrained model. For more details, please refer to section 4.2 in [16].

Opinion Aggregator

To better understand the structures and opinions of the selected answers, we propose an efficient way to aggregate and summarize the answers for a question into a few main opinions through a modified topic model. It was shown that the topical feature vectors produced by the topic models are informative in classification tasks [19,20]. We propose to use a topic model to generate the feature vectors for each answer. However, the number of topics is hard to determine in advance and tuning the number of topics in a topic model is costly because the convergence of a topic model may require considerable iterations. Instead, we propose to aggregate the answers by clustering them according to their topical feature vectors by k-means clustering due to its efficiency in clustering. Each cluster represents an opinion for a question. The meaning of an opinion is represented by the answers in the corresponding cluster.

In our preliminary experiments, we found that the lengths of the answers are substantially different. The traditional topic models—latent Dirichlet allocation [20] and probabilistic latent semantic analysis—are known to suffer from the sparsity problem when the answers are short [21]. Hence, we use the biterm topic model (BTM) [19,22] as an enhanced topic model for distilling the topics that underlie the answers. Rather than using single words, BTM is based on all possible unordered word pairs (ie, biterms) so as to alleviate the sparsity problem caused by applying traditional topic models to short texts. Specifically, we consider all the answers as a mixture of topics, where each biterm is drawn from a specific topic independently. The probability that a biterm is drawn from a specific topic is further captured by the likelihood that both words in the biterm are drawn from the topic.

The estimation and inference based on the BTM (ie, the affiliation of the answers to the topics) can be obtained by Gibbs sampling based on posterior distributions [19]. By applying BTM, a topical feature vector is obtained for each answer. The dimension we used is 40. The topical feature vectors for all answers are clustered by the k-means algorithm, where the best number of the clusters k* is selected using the Silhouette coefficient [23,24]. Each opinion is named by the supportive answer (in the cluster) that is closest to the centroid of a cluster. Its popularity is represented by the number of answers belonging to that cluster. For each opinion, the word frequencies of the corresponding answers are counted. The most frequent words are used together with the opinion name for interpretation.

Proposed Pipeline

Data Set

We conducted experiments on the CORD-19 data set. The data set contains 47,000 articles along with their titles, abstracts, publication dates, and other metadata. Throughout this paper, we use version 2020-04-03 of the data set. In total, six questions from the CORD-19 data set are used as a demonstration. The proposed pipeline is applicable to general questions.

Context Retriever

We concatenated the title and the abstract of an article as a single context passage for calculating the similarity in the context retrieving step. Note that the Natural Language Toolkit' stop words [25] are not included in both the context passages and the questions. When calculating the similarity score with BM25, k=1.5 and b=0.75 were used. After ranking the context passages, we only keep the top 100 as the input of the question-answering model.

Question-Answering Model

The top 100 abstracts selected by the context retriever are used as the input context passages to generate answers to the related question. A valid answer is no longer than 50 words and the end position is always behind the start position. The context passages where the question-answering model fails to generate a valid answer are discarded, and only the valid answers are used in the opinion aggregator.

Opinion Aggregator

The number of topics used in the BTM is 40. The range of candidate numbers of opinions (k) for k-means is from 2 to 5. For each opinion, the words with high frequency are selected as the top words manually.

Results

Principal Results

The results based on all CORD-19 articles, articles published since 2019, and those published since 2020 are summarized in Tables 1-3. For each question, the opinions acquired by the proposed pipeline are presented. Along with each opinion, its corresponding high-frequency words are also listed to facilitate understanding and interpretation.

For the results acquired based on all CORD-19 articles, we further demonstrate the trend of research over time through the popularity of opinions. Figure 2 presents the number of articles under each opinion published before 1990, during 1990-1999, 2000-2009, and 2010-2018, and since 2019 for all questions.

Table 1. Opinions, the number of answers supporting every question, and the words with the highest frequency for each opinion based on all articles.

Question and opinion	Number of answers[a]	Top words in answers
Question 1: What risk factors contribute to the severity of COVID-19?		
1. Transmission and spread	5	Transmission, spread
2. Environmental factors	9	Factors, host, viral, environmental
3. Allergy and preventive therapies	4	N/A[b]
4. Exposure, infection, and clinical disease	56	Disease, factors, viral, clinical, age, inflammatory
5. Interventions for modifiable risk factors	5	Risk, factors
Question 2: How does chronic obstructive pulmonary disease affect patients?		
1. Substantial morbidity	6	Morbidity, mortality, hospital
2. Exacerbations	82	Exacerbations, pulmonary, chronic, obstructive, respiratory, acute
Question 3: What real-time genomic tracking tools exist?		
1. Polymerase chain reaction	12	PCR, RT, real, time, reverse
2. Annotation tools	53	Tracking, genomic, sequencing
3. Microarrays	5	Viral, genome
4. Web-based tools	6	Analysis, genomic
5. Tools to manipulate this system are growing	6	Computational, tracking
Question 4: Which nonpharmaceutical interventions limit transmission?		
1. Quarantine, isolation, and social distancing	72	Quarantine, interventions, social, isolation, distancing, closure
2. Invasive devices	7	Vaccines, antiviral
3. Education about alcohol-based hand sanitizer, and education about hand sanitizer and face masks	8	Hand, face, masks, hygiene, washing, sanitizer, alcohol, based
Question 5: What are the most important barriers to compliance?		
1. Physical barriers	14	Species, physical
2. Perceived barriers	74	Control, lack, perceived, infection, knowledge
Question 6: How is artificial intelligence being used in real-time health delivery?		
1. Text mining	5	Remote, monitoring, control
2. Data fusion	65	Public, epidemic, decision, surveillance, detection
3. Monitors and disseminates online information about emerging infectious diseases	4	Online, information
4. Reactive. It is necessary to understand the complexity and interactions of integrated environmental health risks	4	Necessary
5. Can be incorporated into any real-time polymerase chain reaction assay	4	Incorporated, RT, PCR, assay

[a]Number of answers classified to each opinion (ie, in each k-means cluster).

[b]N/A: not applicable.

Table 2. Opinions, the number of answers supporting every question, and the words with the highest frequency for each opinion in articles published since 2019.

Question and opinion	Number of answers[a]	Top words in answers
Question 1: What risk factors contribute to the severity of COVID-19?		
1. Global spread	6	Transmission, spread
2. Genetic and environmental factors	6	Environmental, genetic
3. Older age and high number of comorbidities	56	Age, patients, infections, viral, diabetes, acute, hypertension, illness, older, comorbidities, respiratory, pathogens
4. Interventions and intensity	5	Interventions, potential
Question 2: How does chronic obstructive pulmonary disease affect patients?		
1. May affect the disease status	63	Disease, chronic, pulmonary, clinical, immune, virus, obstructive, cells, infections, asthma
2. According to the severity of the disease and patients were followed up to the clinical endpoint	5	Treatment, severity, disease
3. Treated with antibiotics	6	Higher, risk
4. They will also be affected as ordinary citizens	6	Mortality, morbidity
Question 3: What real-time genomic tracking tools exist?		
1. Benchmark databases	8	Models, multiple
2. Real-time reverse transcription polymerase chain reaction and genomic sequencing	12	PCR, quantitative, sequencing, RT, reverse, transcription, TaqMan
3. Close contacts	54	Analysis, tools, hosts, sequence, polymerase, genomic
4. Extensive testing and case tracking	5	Testing
5. Prediction tools with distinct algorithms	5	Prediction, forecasting
Question 4: Which nonpharmaceutical interventions limit transmission?		
1. Protective measures and social distancing	10	Social, distancing, measures, quarantine
2. Limit further transmission	57	Measures, epidemic, social, host, infection, travel, reduce, screening, testing, population, transmission, limit, viral, vascular, ml
3. Continuum limit of this model is a system of delay differential equations	4	N/A[b]
4. No spurious polymerase chain reaction amplification occurred from the genomic RNA or DNA of other pathogens	3	PCR, amplification, RNA
5. Without other interventions	11	Health, public, China, implemented
Question 5: What are the most important barriers to compliance?		
1. Species barriers	11	Species, fitness
2. Extracellular barriers	76	Hand, compliance, hygiene, lack, health, epithelial, practices, respirators, information, medical
Question 6: How is artificial intelligence being used in real-time health delivery?		
1. Real-world study	72	Data, based, epidemic, system, model, remote, decision, information, provide, tools, health, control, detection, infectious, surveillance, CRISPR
2. Real-time reverse transcription polymerase chain reaction was used to detect the new coronavirus in respiratory samples	9	PCR, RT, swab

[a]Number of answers classified to each opinion (ie, in each k-means cluster).

[b]N/A: not applicable.

Table 3. Opinions, the number of answers supporting every question, and the words with the highest frequency for each opinion in the articles published since 2020.

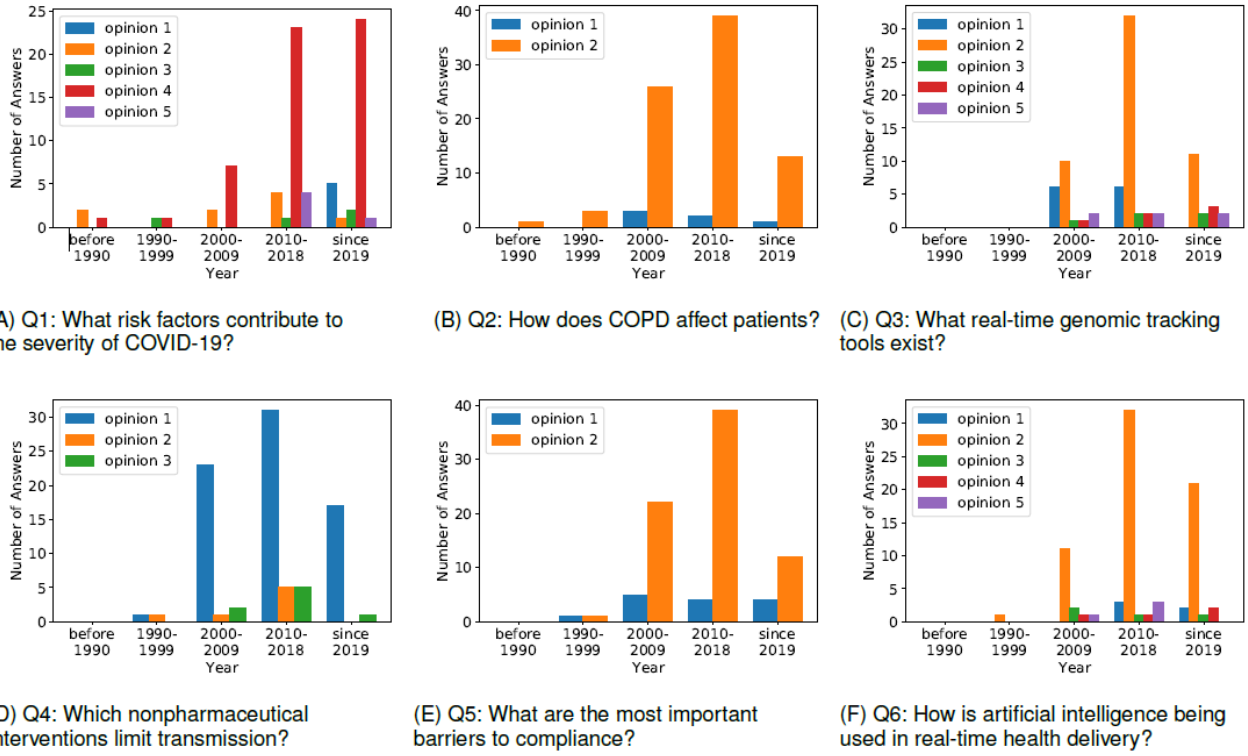
Question and opinion	Number of answers[a]	Top words in answers
Question 1: What risk factors contribute to the severity of COVID-19?		
1. Transmission and spread	5	Transmission, spread, characteristics
2. Nosocomial infection	58	Age, cases, patients, higher, risk, incidence, older, respiratory, diabetes, hypertension, comorbidities, severe, pneumonia, admission, acute, infection, development, lower, viral, levels, elevated, conditions
3. Natural and social factors	7	Interventions, natural
4. Knowledge of the confirmed case fatality risk	4	Screening, knowledge, cCFR
Question 2: How does chronic obstructive pulmonary disease affect patients?		
1. It affects all age groups, including newborns and older adults	4	Increased, all, age, groups
2. According to the severity of the disease, patients were followed up to the clinical endpoint	4	Disease, severity, followed, clinical
3. They will also be affected as ordinary citizens	3	Affected
4. Mixing of viral components during coinfection alters pathogenic outcomes	5	Injury, higher, function
5. Coronavirus pneumonia	57	Mortality, clinical, changes, host, respiratory, high, growth, chronic, disease, asthma, immune, viral, reduced, cases, symptomatic, virus, effects
Question 3: What real-time genomic tracking tools exist?		
1. CodSeqGen	5	Models, multiple
2. Quantitative real-time polymerase chain reaction assays	7	PCR, rt, quantitative, assays
3. Tools and analytical capabilities	9	Analytical, epidemiological, twitter, tracing
4. Sequenced	5	Based
5. Close contacts	53	Data, testing, human, analysis, reverse, methods
Question 4: Which nonpharmaceutical interventions limit transmission?		
1. Social distancing and movement restrictions	11	Social, distancing, restrictions, measures, travel, movement
2. Multidisciplinary comprehensive interventions	6	N/A[b]
3. Screening program	5	Screening
4. Public health intervention	12	Health, public, China, measures, implemented, reduce
5. Limit further transmission	49	Measures, epidemic, social, pharmaceutical, infection, outbreaks, limit, population, vascular, ml, distancing, human, rate, reduce, peak, prevention, coverage, closures, testing, importation, respond
Question 5: What are the most important barriers to compliance?		
1. Infectious pathogens	80	Transmission, virus, medical, protection, research, cross, practice, measures, patients, immune, respiratory, ophthalmic, infection, staff, wildlife, limited, disinfection, isolation
2. Population movement	7	Cellular
Question 6: How is artificial intelligence being used in real-time health delivery?		
1. Cesarean section	5	Monitor
2. Modified mRNA (modRNA)	63	Data, rt, PCR, clinical, epidemic, system, model, modified

Question and opinion	Number of answers[a]	Top words in answers
3. A combination of repeated swab tests and computed tomography scanning	4	Repeated
4. Methods and findings using the traditional SEIR model	5	World, model, network
5. Drug delivery	6	Surveillance, drug, systems

[a]Number of answers classified to each opinion (ie, in each k-means cluster).

[b]N/A: not applicable.

Figure 2. The graphs show the number of articles affiliated with each opinion during different periods for all questions. The numbers are based on all articles. The interpretation of the opinions is shown in Table 1. COPD: chronic obstructive pulmonary disease.



An Example to Illustrate the NLP Pipeline

We used the question "What are the most important barriers to compliance?" and the articles published since 2019 as an example to illustrate the proposed pipeline. First, the top relevant passages are selected by the context retriever. For example, one relevant passage selected is as follows:

> *Timely detection of novel coronavirus (2019-nCoV) infection cases is crucial to interrupt the spread of this virus. We assessed the required expertise and capacity for molecular detection of 2019-nCoV in specialized laboratories in 30 European Union European Economic Area (EU EEA) countries. Thirty-eight laboratories in 24 EU EEA countries had diagnostic tests available by 29 January 2020. A coverage of all EU EEA countries was expected by mid-February. Availability of primers/probes, positive controls and personnel were main implementation barriers.*

Second, the question and the passage are combined as a single sequence as an input to the BERT model. According to the common BERT practice, the sequence starts with the special <CLS> token, and the question and passage are separated by the <SEP> token. The pretrained BERT question-answering model returns the start (Availability) and end locations (personnel), and the resulting answer is a fraction of the passage "Availability of primers/probes, positive controls and personnel." The other passages returned by the BM25 context retriever were also used as input of the pretrained BERT question-answering model, and the answer in each passage is identified similarly.

Third, the answers in the passages are used as the input of the BTM model to obtain feature vectors, which are the input of the k-means clustering used to estimate the number of clusters (ie, opinions in response to the COVID-19–related questions). For the question concerning barriers to compliance, two opinions (ie, clusters) were identified. The number of answers classified to each opinion, the top words in the answers for each opinion, and some sample answers for each opinion are shown in Table S1 in Multimedia Appendix 1. The results for another question (Which nonpharmaceutical interventions limit transmission?) are shown in Table S2 in Multimedia Appendix 1. The results for all questions given articles from different periods are discussed in the following sections.

Results Based on All CORD-19 Articles

For the first question (What risk factors contribute to the severity of COVID-19?), over half of the answers are related to exposure, infection, and the pre-existence of other clinical diseases. The distributions of opinions during different periods for question 1 are shown in Figure 2A. In the last decade, more studies are related to environmental factors and related interventions. Recently, the number of studies on transmission and spread has increased substantially, which agrees with the significantly higher transmissibility of COVID-19. Regarding the second question (How does chronic obstructive pulmonary disease affect patients?), most studies focus on the process of chronic obstructive pulmonary disease (COPD) compared to morbidity. The number of articles on this opinion has increased gradually over the years, as shown in Figure 2B.

For the third question (What real-time genomic tracking tools exist?), research about real-time genomic tracking tools for coronaviruses started to gain attention after the year 2000. Most efforts have been focused on annotation tools for genomic sequencing and tracking. The second topic frequently studied is related to polymerase chain reaction (PCR). The other three topics, which include novel tracking methods (microarrays, web-based tools, and computational tools), received relatively less attention before 2019, as shown in Figure 2C. However, they have attracted more and more attention during the COVID-19 pandemic, as demonstrated when comparing the numbers in the past year to those in the past 10 years.

For the fourth question (Which nonpharmaceutical interventions limit transmission?), three topics have been studied: (1) quarantine and social distancing, (2) vaccines, and (3) education about using hand sanitizer and face masks. Most publications focus on the first topic. More publications about topics 2 and 3 appeared in the second decade of this century than the first decade, as shown in Figure 2D.

In answering the fifth question (What are the most important barriers to compliance?), most of the studies tend to focus more on perceived barriers than physical barriers, as shown in Figure 2E. For the sixth question (How is artificial intelligence being used in real-time health delivery?), research on using artificial intelligence (AI) has attracted increasing attention since the year 2000, as shown in Figure 2F. Most of the selected publications contribute to using AI for detection, surveillance, epidemic research, and decisions. Other AI applications include remote and online monitoring and bioinformatics applications.

Results Based on CORD-19 Articles Since 2019 and Since 2020

To better understand the studies published near the outbreak of COVID-19, we repeated the analysis using the publications that appeared since 2019 and since 2020, respectively. The results are compared to those based on all publications. Specifically, we compared the opinions of the answers and the top words among these time frames and highlighted research trends.

The recently reported high-risk factors (question 1) are more related to senior people [26,27] and pre-existing medical conditions [28], which have been reported to be important characteristics of COVID-19. Compared to previous publications about the effect of COPD (question 2) that focused on various symptoms and pre-existing conditions, recent publications have studied these together with mortality. The reason may be that patients with COVID-19 experience a wide range of disease severity (eg, asymptomatic, mild disease, severe disease, death). Furthermore, the number of deaths is substantially larger than the number of deaths from other viruses from the same family [29]. Regarding nonpharmaceutical interventions that limit transmission (question 4), the recent emphasis has been on screening and testing, given that COVID-19 is highly contagious and spreads rapidly. The barriers to compliance (question 5) have shifted from perceived barriers to the challenges of COVID-19, such as rapid transmission and the implementation of practical interventions (eg, protection, research, disinfection, and isolation). As indicated by the publications, the most noticeable increase of an AI application (question 6) is related to PCR tests.

Discussion

Principal Findings

To determine what the CORD-19 literature reports about COVID-19–related questions, we proposed a pipeline to select relevant passages with a BM25 context retriever, find answers by applying a BERT question-answering model, and distill the main opinions using a BTM-based opinion aggregator and k-means clustering. The results demonstrate the usefulness of the proposed method in answering COVID-19–related questions with main opinions as well as determining their popularity in the literature and capturing the trends in research about COVID-19 and other relevant strains of coronavirus in recent years. The discovered trends agree with the findings about and challenges related to COVID-19.

Limitations

There are several limitations in our current study. First, in the current analysis, we only incorporated the abstracts of articles as the knowledge source. Considering full articles would provide more information but require more computational power. We may also need to account for multiple answers in an article. Second, for the opinion aggregator, we used a BTM topical model to obtain features for clustering. Semantical paraphrasing methods are expected to result in better representation and clustering results in future work.

Comparison With Prior Work

Compared to several very recent studies using question-answering models for CORD-19, our study has several unique contributions. First, our proposed system can automatically aggregate the answers to a user query into a few main opinions, which can prevent users from getting overwhelmed by a large number of differing answers and help users grasp the current major opinions related to the query.

Second, compared to a large set of independent articles, the aggregated opinions provide a more robust representation of the research ideas, together with their popularity as quantified by the number of supportive articles. Third, considering COVID-19 research is still ongoing, there will likely be new ideas, discoveries, and arguments put forward in the future that

will better answer questions about COVID-19. Compared to arbitrary answers generated by these recently proposed methods or developed systems, our approach shows all the existing opinions, which may be more suitable for open questions and more informative for researchers. Finally, we demonstrate comprehensive data analysis results and interpretations to provide useful insights into COVID-19 research.

Conclusions

Compared with other methods for literature retrieval and answer generation, the opinion aggregation of our method leads to a more interpretable, robust, and comprehensive question-specific literature review. The results demonstrate the usefulness of the proposed method in answering COVID-19–related questions with main opinions and capturing the coronavirus research trends of recent years.

Acknowledgments

The research work of ZH-L and JXW was in part supported by the Department of Biostatistics at St Jude Children's Research Hospital (SJCRH) and the American Lebanese Syrian Associated Charities (ALSAC). JXW's research was also partly supported by the Academic Office Program at SJCRH.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 15 KB - jmir_v23i3e22860_app1.docx]

References

1. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
2. A Chronicle on the SARS Epidemic. Chinese Law & Government 2014 Dec 07;36(4):12-15. [doi: 10.2753/clg0009-4609360412]
3. Zumla A, Hui DS, Perlman S. Middle East respiratory syndrome. The Lancet 2015 Sep 05;386(9997):995-1007. [doi: 10.1016/s0140-6736(15)60454-8] [Medline: 26049252]
4. Lu Wang L, Lo K, Chandrasekhar Y, Reas R, Yang J, Eide D, et al. CORD-19: The Covid-19 Open Research Dataset. ArXiv. Preprint posted online on April 22, 2020. [Medline: 32510522]
5. CORD-19 Search. URL: https://cord19.aws/ [accessed 2020-07-21]
6. tmCOVID-19. URL: https://tmcovid.com/ [accessed 2020-07-22]
7. WellAi - Machine Learning Analytics For Researchers. URL: https://wellai.health/ [accessed 2020-07-22]
8. Neural Covidex. URL: https://covidex.ai/ [accessed 2020-07-23]
9. Tang R, Nogueira R, Zhang E, Gupta N, Cam P, Cho K, et al. Rapidly bootstrapping a question answering dataset for COVID-19. ArXiv. Preprint posted online on April 23, 2020 [FREE Full text]
10. Su D, Xu Y, Yu T, Siddique F, Barezi E, Fung P. CAiRE-COVID: A question answering and multi-document summarization system for COVID-19 research. 2020 Presented at: Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020; December 2020; Online. [doi: 10.18653/v1/2020.nlpcovid19-2.14]
11. Esteva A, Kale A, Paulus R, Hashimoto K, Yin W, Radev D, et al. CO-Search: COVID-19 information retrieval with semantic search, question answering, and abstractive summarization. ArXiv. Preprint posted online on June 17, 2020 [FREE Full text]
12. Lewis M, Liu Y, Goyal N, Ghazvininejad M, Mohamed A, Levy O, et al. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. 2020 Presented at: 58th Annual Meeting of the Association for Computational Linguistics; July 2020; Online p. 7871-7880. [doi: 10.18653/v1/2020.acl-main.703]
13. Chen D, Fisch A, Weston J, Bordes A. Reading Wikipedia to answer open-domain questions. 2017 Presented at: 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); July 2017; Vancouver, Canada p. 1870-1879. [doi: 10.18653/v1/p17-1171]
14. Trotman A, Puurula A, Burgess B. Improvements to BM25 and language models examined. 2014 Presented at: 2014 Australasian Document Computing Symposium; November 2014; Melbourne, Victoria, Australia. [doi: 10.1145/2682862.2682863]
15. Cacheda F, Fernandez D, Novoa FJ, Carneiro V. Early Detection of Depression: Social Network Analysis and Random Forest Techniques. J Med Internet Res 2019 Jun 10;21(6):e12554 [FREE Full text] [doi: 10.2196/12554] [Medline: 31199323]
16. Devlin J, Chang M, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. 2019 Presented at: 2019 Conference of the North American Chapter of the Association for Computational Linguistics:

Human Language Technologies, Volume 1 (Long and Short Papers); June 2019; Minneapolis, MN, USA p. 4171-4186. [doi: 10.18653/v1/N19-1423]
17. Li F, Jin Y, Liu W, Rawat BPS, Cai P, Yu H. Fine-Tuning Bidirectional Encoder Representations From Transformers (BERT)-Based Models on Large-Scale Electronic Health Record Notes: An Empirical Study. JMIR Med Inform 2019 Sep 12;7(3):e14830 [FREE Full text] [doi: 10.2196/14830] [Medline: 31516126]
18. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. HuggingFace's transformers: State-of-the-art natural language processing. 2020 Presented at: 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations; October 2020; Online p. 38-45. [doi: 10.18653/v1/2020.emnlp-demos.6]
19. Yan X, Guo J, Lan Y, Cheng X. A biterm topic model for short texts. 2013 Presented at: 22nd International Conference on World Wide Web; May 2013; Rio de Janeiro, Brazil p. 1445-1456. [doi: 10.1145/2488388.2488514]
20. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. Journal of Machine Learning Research 2003;3(Jan):993-1022 [FREE Full text]
21. Hong L, Davison BD. Empirical study of topic modeling in Twitter. 2010 Presented at: First Workshop on Social Media Analytics; July 2010; Washington, DC, USA p. 80-88. [doi: 10.1145/1964858.1964870]
22. Yamada Y, Shinkawa K, Shimmei K. Atypical Repetition in Daily Conversation on Different Days for Detecting Alzheimer Disease: Evaluation of Phone-Call Data From Regular Monitoring Service. JMIR Ment Health 2020 Jan 12;7(1):e16790 [FREE Full text] [doi: 10.2196/16790] [Medline: 31934870]
23. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 1987 Nov;20:53-65. [doi: 10.1016/0377-0427(87)90125-7]
24. Kaufman L, Rousseeuw P. Finding Groups in Data: An Introduction to Cluster Analysis. New York, NY, USA: John Wiley & Sons; 2009.
25. Bird S, Klein E, Loper E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. Sebastopol, CA, USA: O'Reilly Media Inc; 2009.
26. Centers for Disease Control and Prevention. Older Adults and COVID-19. URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html [accessed 2021-02-17]
27. Koff WC, Williams MA. Covid-19 and Immunity in Aging Populations — A New Research Agenda. N Engl J Med 2020 Aug 27;383(9):804-805. [doi: 10.1056/nejmp2006761]
28. Centers for Disease Control and Prevention. Certain Medical Conditions and Risk for Severe COVID-19 Illness. URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html [accessed 2021-02-17]
29. World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report – 46. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf?sfvrsn=96b04adf_4 [accessed 2021-02-10]

Abbreviations

AI: artificial intelligence
BERT: bidirectional encoder representations from transformers
BTM: biterm topic model
COPD: chronic obstructive pulmonary disease
CORD-19: COVID-19 Open Research Dataset Challenge
MERS-CoV: Middle East respiratory syndrome coronavirus
PCR: polymerase chain reaction
SARS-CoV: severe acute respiratory syndrome coronavirus
TF-IDF: term frequency–inverse document frequency

Edited by G Eysenbach, C Basch; submitted 25.07.20; peer-reviewed by S Zheng, S Doan; comments to author 20.09.20; revised version received 16.11.20; accepted 31.01.21; published 19.03.21.

Please cite as:
Lu ZH, Wang JX, Li X
Revealing Opinions for COVID-19 Questions Using a Context Retriever, Opinion Aggregator, and Question-Answering Model: Model Development Study
J Med Internet Res 2021;23(3):e22860
URL: https://www.jmir.org/2021/3/e22860
doi:10.2196/22860
PMID:33739287

Original Paper

Insights From the SmokeFree.gov Initiative Regarding the Use of Smoking Cessation Digital Platforms During the COVID-19 Pandemic: Cross-sectional Trends Analysis Study

Sherine El-Toukhy[1], MA, PhD

Division of Intramural Research, National Institute on Minority Health & Health Disparities, National Institutes of Health, Bethesda, MD, United States

Corresponding Author:
Sherine El-Toukhy, MA, PhD
Division of Intramural Research
National Institute on Minority Health & Health Disparities
National Institutes of Health
7201 Wisconsin Ave
Bethesda, MD, 20814
United States
Phone: 1 3015944743
Email: sherine.el-toukhy@nih.gov

Abstract

Background: Smoking is a plausible risk factor for COVID-19 progression and complications. Smoking cessation digital platforms transcend pandemic-driven social distancing and lockdown measures in terms of assisting smokers in their quit attempts.

Objective: This study aims to examine trends in the number of visitors, followers, and subscribers on smoking cessation digital platforms from January to April 2020 and to compare these traffic data to those observed during the same 4-month period in 2019. The examination of prepandemic and postpandemic trends in smoking cessation digital platform traffic can reveal whether interest in smoking cessation among smokers is attributable to the COVID-19 pandemic.

Methods: We obtained cross-sectional data from daily visitors on the SmokeFree website; the followers of six SmokeFree social media accounts; and subscribers to the SmokeFree SMS text messaging and mobile app interventions of the National Cancer Institute's SmokeFree.gov initiative platforms, which are publicly available to US smokers. Average daily percentage changes (ADPCs) were used to measure trends for the entire 2020 and 2019 study periods, whereas daily percentage changes (DPCs) were used to measure trends for each time segment of change within each 4-month period. Data analysis was conducted in May and June 2020.

Results: The number of new daily visitors on the SmokeFree website (between days 39 and 44: DPC=18.79%; 95% CI 5.16% to 34.19%) and subscribers to the adult-focused interventions QuitGuide (between days 11 and 62: DPC=1.11%; 95% CI 0.80% to 1.43%) and SmokeFreeTXT (between days 11 and 89: DPC=0.23%; 95% CI 0.004% to 0.47%) increased, but this was followed by declines in traffic. No comparable peaks were observed in 2019. The number of new daily subscribers to quitSTART (ie, the teen-focused intervention) trended downward in 2020 (ADPC=−1.02%; 95% CI −1.88% to −0.15%), whereas the overall trend in the number of subscribers in 2019 was insignificant (P=.07). The number of SmokeFree social media account followers steadily increased by <0.1% over the 4-month study periods in 2019 and 2020.

Conclusions: Peaks in traffic on the SmokeFree website and adult-focused intervention platforms in 2020 could be attributed to an increased interest in smoking cessation among smokers during the COVID-19 pandemic. Coordinated campaigns, especially those for adolescents, should emphasize the importance of smoking cessation as a preventive measure against SARS-CoV-2 infection and raise awareness of digital smoking cessation platforms to capitalize on smokers' heightened interest during the pandemic.

(J Med Internet Res 2021;23(3):e24593) doi:10.2196/24593

KEYWORDS

COVID-19; smoking; cessation; mHealth; risk; digital platform; social distancing; lockdown; trend

Introduction

The novel SARS-CoV-2 causes COVID-19. Initially, COVID-19 was identified in December 2019 from a cluster of pneumonia cases of unknown causes in Wuhan, China [1]. COVID-19 has spread throughout the globe and has claimed hundreds of thousands of lives. In the United States, the official number of confirmed cases has surpassed 20 million, and 345,000 fatalities have been reported as of December 31, 2020 [2].

As researchers have attempted to identify risk factors that are associated with SARS-CoV-2 infection and COVID-19–related complications, smoking has emerged as a plausible candidate [3]. Smoking is detrimental to health. Therefore, smokers are more susceptible to the development of disease (specifically pulmonary illnesses) and poor symptoms and outcomes (ie, those associated with health conditions such as asthma and obesity) [4]. Since COVID-19 is primarily a respiratory illness, it is important to note that smoking-induced pathophysiological changes can result in weakened immune responses, inflammation marker development, genetic changes in lung tissue, structural changes in the respiratory tract, and the dysfunction of the lungs. Therefore, smokers are more susceptible to the onset of respiratory illnesses [5,6]. Indeed, studies have shown that smokers have a higher risk of bacterial and viral infection (eg, invasive pneumococcal infection, influenza infection, and infection from coronaviruses like Middle East respiratory syndrome–related coronavirus) than nonsmokers [6,7]. Systematic reviews and meta-analytic evidence have shown that smoking is a risk factor for coronavirus infection, severity, and mortality [8-12]. However, evidence has been inconclusive and contradictory [13,14].

Smoking cessation is an advisable preventive measure against SARS-CoV-2 infection [15]. Compared to nonsmokers, smokers and ex-smokers have significantly higher stress levels due to their susceptibility to SARS-CoV-2 infection and COVID-19 severity [16], which are known health belief model constructs that affect behavioral intentions and behaviors [17]. The COVID-19 pandemic can be considered a naturally occurring cueing event that has motivated smokers to engage in preventive behaviors for reducing their risk of contracting COVID-19 [18]. For example, studies have shown that receiving a new disease diagnosis is associated with quitting smoking [19,20]. Indeed, public health professionals have advised smokers to use interventions that are proven to be effective in helping smokers quit smoking, which can reduce their risk of SARS-CoV-2 infection and COVID-19–related complications [21-23].

As a result of the lockdown and physical distancing measures that have been implemented to reduce the transmission of SARS-CoV-2, digital platforms for smoking cessation resources and interventions have superseded in-person cessation support services [24,25]. Web searches and social media postings provide insights into public opinions about health issues and opportunities for surveilling disease outbreaks and symptoms [26,27]. Furthermore, digital interventions have proven to be effective in inducing behavioral changes, such as quitting smoking [28]. With regard to COVID-19, a study on Google Trends reported that there was no increase in the incidence of smoking cessation–related search terms (eg, “smoking cessation,” “nicotine gum,” and “quit smoking”) from January to April 2020 [29]. Furthermore, hashtags such as #Quit4COVID are prominent on Twitter [30], and researchers have used Facebook data to index social connectedness data that correlate with the geographic spread of COVID-19 [31]. Preliminary evidence from the United Kingdom has also shown that there has been no increase in the number of downloads for Smoke Free (ie, a smoking cessation mobile app) as a result of the COVID-19 outbreak [32].

Traffic on smoking cessation digital platforms is a proxy measure of smokers’ interest in smoking cessation and quit attempts during the COVID-19 pandemic. One such suite of digital platforms that is freely available to smokers in the United States is the National Cancer Institute’s SmokeFree.gov initiative (SFGI) [33,34]. Initially, the SFGI had a single website that launched in 2003 [35]. The SFGI now has multiple platforms, including the main SFGI website, 6 social media accounts, 2 mobile apps (ie, QuitGuide and quitSTART), and 6 SMS text messaging interventions. Around 7-8 million smokers use SFGI platforms every year, and these platforms have an estimated efficacy that ranges from approximately 10% to 30% [35]. The SmokeFree website hosts information on all SFGI smoking cessation programs. SFGI programs typically appear as the top search results on search engines. They are also featured on the webpages of authorities on tobacco control, such as the Centers for Disease Control and Prevention [36] and the Food and Drug administration [37]. SmokeFree SMS text messaging and mobile app interventions for smoking cessation are based on social behavior theory; preliminary studies have shown the acceptability and effectiveness of several SFGI programs [38,39]. The aims of this study are (1) to characterize daily trends in traffic on SFGI digital platforms over 4 months (ie, from January 1 to April 30, 2020); and (2) to compare these trends to those from January to April 2019. We hypothesized that traffic on SFGI digital platforms would increase over the course of the pandemic and that these trends would be qualitatively different from prepandemic trends (ie, those observed in 2019).

Methods

Data were obtained from the National Cancer Institute’s SmokeFree smoking cessation digital platforms [33]. We obtained aggregated data on the number of new daily visitors on the SmokeFree website, the number of followers on six social media platforms on each given day (ie, SmokeFree Veterans Facebook, SmokeFree Women Facebook, SmokeFreeUS Facebook, SmokeFreeUS Instagram, SmokeFreeUS Pinterest, and SmokeFreeUS Twitter), and the number of new daily subscribers to three smoking cessation interventions (ie, the quitSTART mobile app, QuitGuide mobile app, and SmokeFreeTXT). The study period was January to April 2020. We also obtained data from January to April 2019. Data analysis was conducted in May and June 2020. Since the data were deidentified, approval from an institutional review board was not required.

A joinpoint regression analysis [40,41] was used to model trends in the daily traffic on SmokeFree digital platforms over 4 months and any significant changes in individual time segments within the 4-month period. Joinpoint software (National Cancer Institute) has been used to characterize trends in cancer morbidity, cancer mortality, and the prevalence of risky and healthy behaviors. Joinpoint connects several different line segments with different slopes on a log scale at "joinpoints" and identifies the number and time points in which the trend significantly changes. Joinpoint regression is particularly flexible in terms of modeling nonlinear trends with nontraditional curves, including trends with abrupt changes. Such trends may be present in the traffic on digital platforms for smoking cessation during a pandemic.

Joinpoint analysis involves the calculation of annual percent change and average annual percent change values. However, for this study, time was measured in days. Accordingly, the terms "daily percent change" (DPC) and "average daily percent change" (ADPC) were used (a deviation from joinpoint terminology) to reflect DPCs and ADPCs in the number of daily visitors, followers, or subscribers on each digital platform. The DPC characterizes the trend for each time segment in the joinpoint models. The ADPC is a summary measure of the trend for a prespecified fixed interval; it yields a single number that describes the ADPC for a period of time. For this study, ADPC was estimated for the entire 121- and 120-day periods of interest (ie, from January 1 to April 30, 2020 and 2019, respectively). When the ADPC lies within a single joinpoint segment, the ADPC is equal to the DPC for that segment.

For each platform, we specified a joinpoint model with a minimum of 0 joinpoints and a maximum of 5 joinpoints, as per the recommendations for models with ≥27 data points [42]. We arranged to have at least 5 observations from a joinpoint to either end of the data, excluding the first or last joinpoint if it falls on an observation data point. Further, we arranged to have at least 4 observations between two joinpoints, excluding any joinpoint if it falls on an observation data point. We used the grid search method and Bayesian Information Criteria 3 method to identify the best fitting models and select the final model (Multimedia Appendix 1).

This study only involved the use of deidentified data. Therefore, this study is considered "not human subjects research" and does not require institutional review board review or approval, as per the National Institutes of Health policy and section 45, part 46 of the Code of Federal Regulations. Furthermore, since this study is considered "not human subjects research," informed consent is not required.

Results

SmokeFree Website and Social Media Platforms

SmokeFree Website

From January to April 2020, the number of new daily visitors on the SmokeFree website ranged from 5344 to 23,959 (Figure 1). The number of new daily visitors increased by a DPC of 18.79% (95% CI 5.16% to 34.19%) during days 39-44, followed by a slight decrease of <1% (DPC=−0.71%; 95% CI −1.01% to −0.41%) during days 44-91 (Table 1). However, the overall trend from January to April was not significant (ADPC=0.57%; 95% CI −0.07% to 1.23%; *P*=.08; Table 2). In 2019, the number of new daily visitors on the SmokeFree website ranged from 9516 to 18,807. The number of new daily visitors on the website steadily increased by 0.11% (95% CI 0.03% to 0.19%).

Figure 1. The number of new SmokeFree.gov website visitors from January to April 2019 and 2020. The vertical line represents January 20. On this day in 2020, the first laboratory-confirmed COVID-19 case was identified in the United States. This was reported to the Centers for Disease Control and Prevention on January 22, 2020 [43]. Data on model selection appear in Multimedia Appendix 1. DPC: daily percent change. *The DPC is significantly different from 0 at an α level of .05.

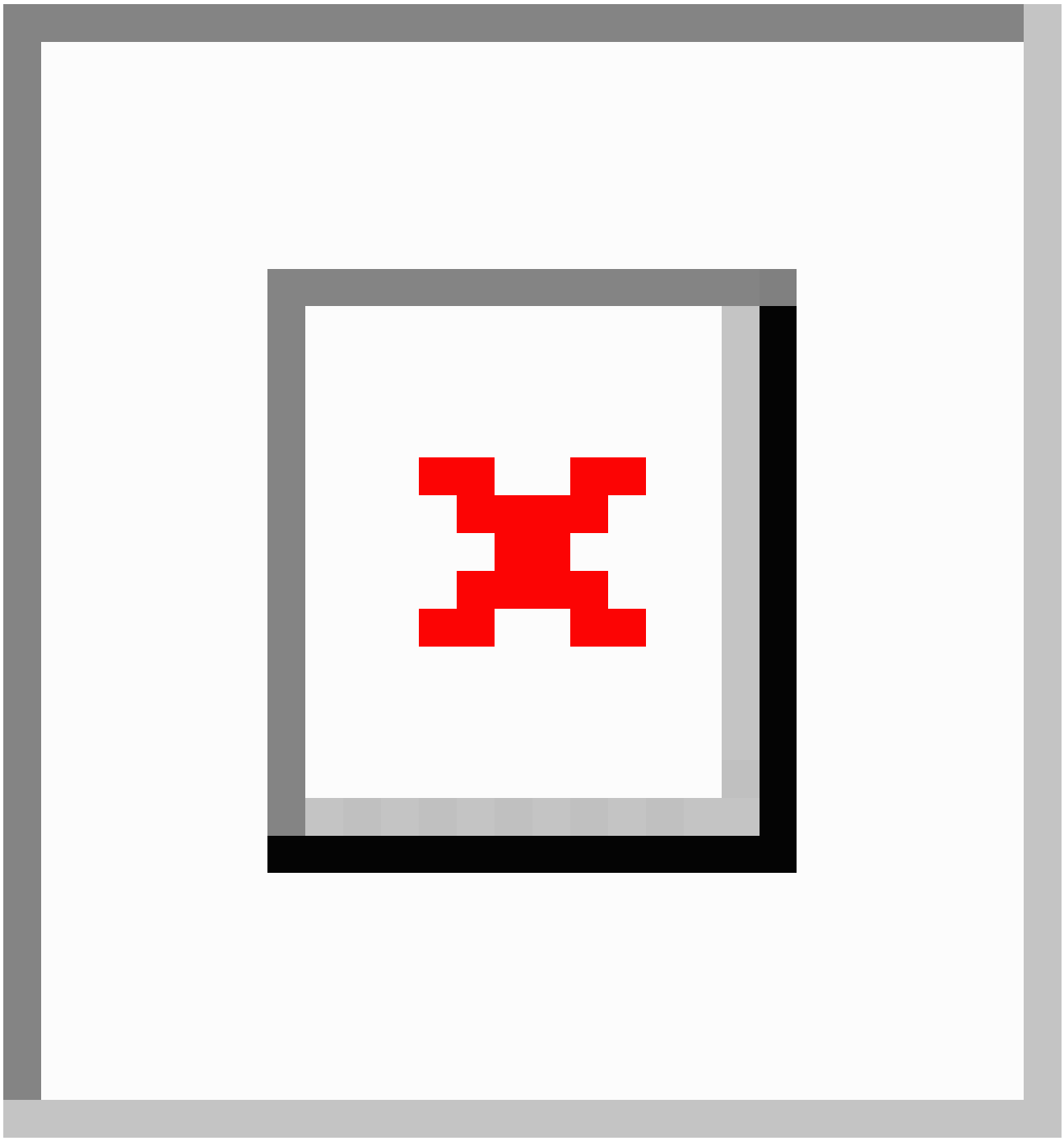


Table 1. Daily percent changes (DPCs) in the number of SmokeFree.gov initiative digital platform visitors, followers, and subscribers from January to April 2019 and 2020.

Digital platform/intervention and time segment	2019[a]			2020[b]		
	Start day	End day	DPC (95% CI)	Start day	End day	DPC (95% CI)
SmokeFree website						
1	1	120	0.116[c] (0.035 to 0.197)	1	39	−0.336 (−0.737 to 0.066)
2	N/A[d]	N/A	N/A	39	44	18.796[c] (5.166 to 34.192)
3	N/A	N/A	N/A	44	91	−0.712[c] (−1.012 to −0.411)
4	N/A	N/A	N/A	91	99	4.959 (−0.308 to 10.506)
5	N/A	N/A	N/A	99	121	−0.425 (−1.333 to 0.490)
SmokeFree Veterans Facebook						
1	1	17	0.047[c] (0.036 to 0.058)	1	21	0.038[c] (0.029 to 0.046)
2	17	22	−0.064 (−0.156 to 0.027)	21	26	0.178[c] (0.081 to 0.274)
3	22	45	0.026[c] (0.020 to 0.033)	26	99	0.021[c] (0.020 to 0.022)
4	45	57	−0.022[c] (−0.042 to −0.003)	99	116	0.097[c] (0.086 to 0.109)
5	57	69	0.072[c] (0.052 to 0.092)	116	121	−0.043 (−0.111 to 0.025)
6	69	120	0.023[c] (0.021 to 0.025)	N/A	N/A	N/A
SmokeFree Women Facebook						
1	1	30	0.014[c] (0.014 to 0.015)	1	7	0.087[c] (0.079 to 0.096)
2	30	52	0.010[c] (0.009 to 0.010)	7	14	0.050[c] (0.041 to 0.058)
3	52	67	0.006[c] (0.005 to 0.008)	14	73	0.038[c] (0.038 to 0.038)
4	67	104	0.017[c] (0.016 to 0.017)	73	98	0.024[c] (0.023 to 0.025)
5	104	111	0.005[c] (0.00009 to 0.011)	98	107	0.040[c] (0.035 to 0.046)
6	111	120	0.024[c] (0.021 to 0.027)	107	121	0.021[c] (0.019 to 0.024)
SmokeFreeUS Facebook						
1	1	9	0.058[c] (0.051 to 0.065)	1	6	0.083[c] (0.065 to 0.102)
2	9	23	0.038[c] (0.034 to 0.041)	6	30	0.038[c] (0.036 to 0.040)
3	23	51	0.022[c] (0.021 to 0.023)	30	36	0.019[c] (0.001 to 0.038)
4	51	95	0.034[c] (0.033 to 0.034)	36	62	0.044[c] (0.042 to 0.046)
5	95	100	0.525[c] (0.031 to 0.073)	62	89	0.024[c] (0.022 to 0.025)
6	100	120	0.036[c] (0.034 to 0.037)	89	121	0.044[c] (0.043 to 0.045)
SmokeFreeUS Instagram						
1	9[e]	17	0.096[c] (0.072 to 0.119)	1	52	0.054[c] (0.052 to 0.055)
2	17	48	0.057[c] (0.054 to 0.060)	52	69	0.141[c] (0.132 to 0.150)
3	48	70	0.032[c] (0.027 to 0.038)	69	75	0.049 (−0.001 to 0.099)
4	70	89	0.053[c] (0.046 to 0.060)	75	108	0.122[c] (0.119 to 0.125)
5	89	105	0.096[c] (0.086 to 0.105)	108	121	0.088[c] (0.076 to 0.100)
6	105	120	0.033[c] (0.023 to 0.042)	N/A	N/A	N/A
SmokeFreeUS Pinterest						

Digital platform/intervention and time segment	2019[a]			2020[b]		
	Start day	End day	DPC (95% CI)	Start day	End day	DPC (95% CI)
1	1	23	0.139[c] (0.128 to 0.149)	1	8	0.064[c] (0.040 to 0.087)
2	23	31	−0.002 (−0.058 to 0.054)	8	59	0.017[c] (0.016 to 0.019)
3	31	51	0.104[c] (0.092 to 0.117)	59	72	0.061[c] (0.050 to 0.071)
4	51	56	0.222[c] (0.087 to 0.356)	72	86	0.006 (−0.003 to 0.015)
5	56	93	0.094[c] (0.089 to 0.099)	86	94	0.065[c] (0.042 to 0.089)
6	93	120	0.044[c] (0.037 to 0.052)	94	121	0.025[c] (0.022 to 0.028)
SmokeFreeUS Twitter						
1	1	27	0.001[c] (0.0009 to 0.002)	1	24	0.008[c] (0.007 to 0.009)
2	27	32	−0.033[c] (−0.046 to −0.020)	24	68	0.002[c] (0.001 to 0.002)
3	32	47	−0.0003 (−0.002 to 0.001)	68	86	−0.007[c] (−0.009 to −0.005)
4	47	120	0.005[c] (0.005 to 0.005)	86	100	0.006[c] (0.004 to 0.009)
5	N/A	N/A	N/A	100	112	−0.009[c] (−0.013 to −0.006)
6	N/A	N/A	N/A	112	121	0.014[c] (0.009 to 0.019)
quitSTART						
1	1	73	−0.333[c] (−0.540 to −0.126)	1	11	−8.825[c] (−12.825 to −4.641)
2	73	120	0.969[c] (0.572 to 1.368)	11	79	−0.223 (−0.480 to 0.033)
3	N/A	N/A	N/A	79	84	12.487 (−6.252 to 34.972)
4	N/A	N/A	N/A	84	121	−2.000[c] (−2.613 to −1.383)
QuitGuide						
1	1	25	−2.100[c] (−3.103 to −1.085)	1	11	−7.457[c] (−10.641 to −4.158)
2	25	120	0.034 (-0.096 to 0.165)	11	62	1.118[c] (0.803 to 1.434)
3	N/A	N/A	N/A	62	80	−3.145[c] (−4.658 to −1.607)
4	N/A	N/A	N/A	80	85	10.136 (−4.467 to 26.972)
5	N/A	N/A	N/A	85	121	−0.338 (−0.845, 0.171)
SmokeFreeTXT						
1	1	120	0.004 (−0.142 to 0.150)	1	11	−5.411[c] (−10.059 to −0.523)
2	N/A	N/A	N/A	11	89	0.235[c] (0.0004 to 0.470)
3	N/A	N/A	N/A	89	94	13.050 (−7.871 to 38.724)
4	N/A	N/A	N/A	94	121	−1.993[c] (−3.095 to −0.879)

[a]In 2019, January includes days 1-31, February includes days 32-59, March includes days 60-90, and April includes days 91-120.

[b]In 2020, January includes days 1-31, February includes days 32-60, March includes days 61-91, and April includes days 92-121.

[c]The DPC is significantly different from 0 at an α level of .05.

[d]N/A: not applicable.

[e]Data on the number SmokeFree Instagram followers were missing for the first 8 days in January 2019.

Table 2. Average daily percent changes (ADPCs) in the number of SmokeFree.gov initiative digital platform visitors, followers, and subscribers from January to April 2019 and 2020.

Digital platform/intervention	ADPC (95% CI) in 2019	ADPC (95% CI) in 2020
SmokeFree website	0.116[a] (0.035 to 0.197)	0.577 (−0.075 to 1.234)
SmokeFree Veterans Facebook	0.023[a] (0.018 to 0.028)	0.038[a] (0.033 to 0.044)
SmokeFree Women Facebook	0.013[a] (0.013 to 0.014)	0.037[a] (0.036 to 0.037)
SmokeFreeUS Facebook	0.034[a] (0.033 to 0.035)	0.039[a] (0.037 to 0.040)
SmokeFreeUS Instagram	0.056[a] (0.053 to 0.059)	0.088[a] (0.085 to 0.091)
SmokeFreeUS Pinterest	0.092[a] (0.084 to 0.099)	0.028[a] (0.026 to 0.031)
SmokeFreeUS Twitter	0.002[a] (0.001 to 0.002)	0.002[a] (0.001 to 0.002)
quitSTART	0.179 (−0.019 to 0.377)	−1.024[a] (−1.882 to −0.159)
QuitGuide	−0.399[a] (−0.628 to −0.170)	−0.351 (−1.068 to 0.370)
SmokeFreeTXT	0.004 (−0.142 to 0.150)	−0.252 (−1.229 to 0.734)

[a]Indicates that the ADPC is significantly different from 0 at an α level of .05.

Social Media Platforms

In 2020, SmokeFree social media accounts had an average number of daily followers that ranged from 605 (SmokeFree Veterans Facebook) to 31,623 (SmokeFree Women Facebook; Figures 2 and 3). Social media accounts exhibited a steady but small increase in the number of daily followers (<0.1%); the ADPCs ranged from a low of 0.002% (SmokeFree US Twitter: 95% CI 0.001% to 0.002%) to a high of 0.08% (SmokeFreeUS Instagram: 95% CI 0.08% to 0.09%). In 2019, the average number of daily followers on SmokeFree social media accounts ranged from 549 (SmokeFree Veterans Facebook) to 28,244 (SmokeFree Women Facebook). SmokeFree social media accounts exhibited a steady but small increase in the number of daily followers (≤0.1%) from January to April 2019; the ADPCs ranged from a low of 0.002% (SmokeFreeUS Twitter 95% CI 0.001% to 0.002%) to a high of 0.09% (SmokeFreeUS Pinterest: 95% CI 0.08% to 0.09%).

Figure 2. The number of SmokeFree VET Facebook, SmokeFree Women Facebook, and SmokeFreeUS Facebook followers from January to April 2019 and 2020. The vertical line represents January 20. On this day in 2020, the first laboratory-confirmed COVID-19 case was identified in the United States. This was reported to the Centers for Disease Control and Prevention on January 22, 2020 [43]. Data on model selection appear in Multimedia Appendix 1. DPC: daily percent change; VET: Veterans. *The DPC is significantly different from 0 at an α level of .05.

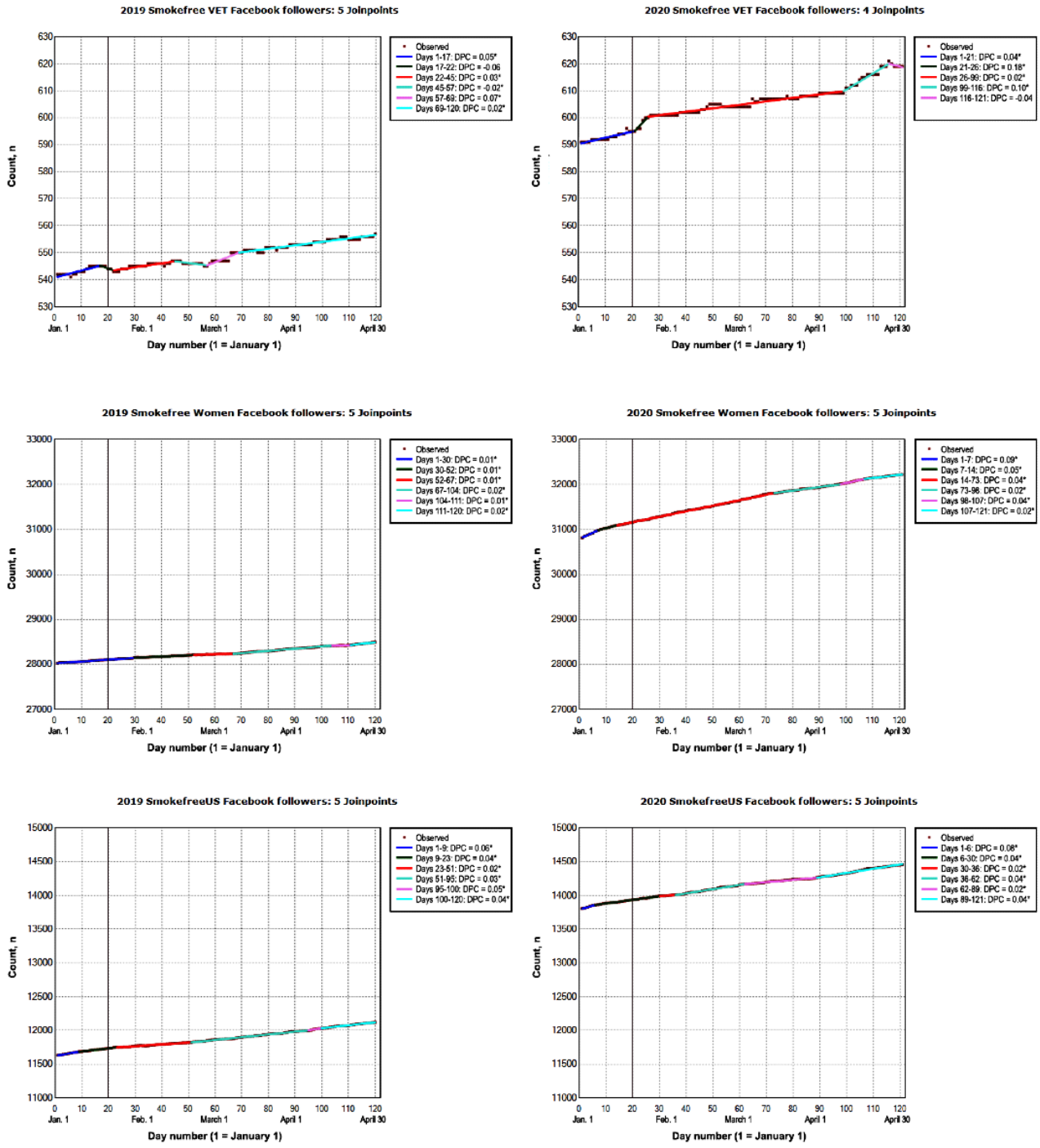
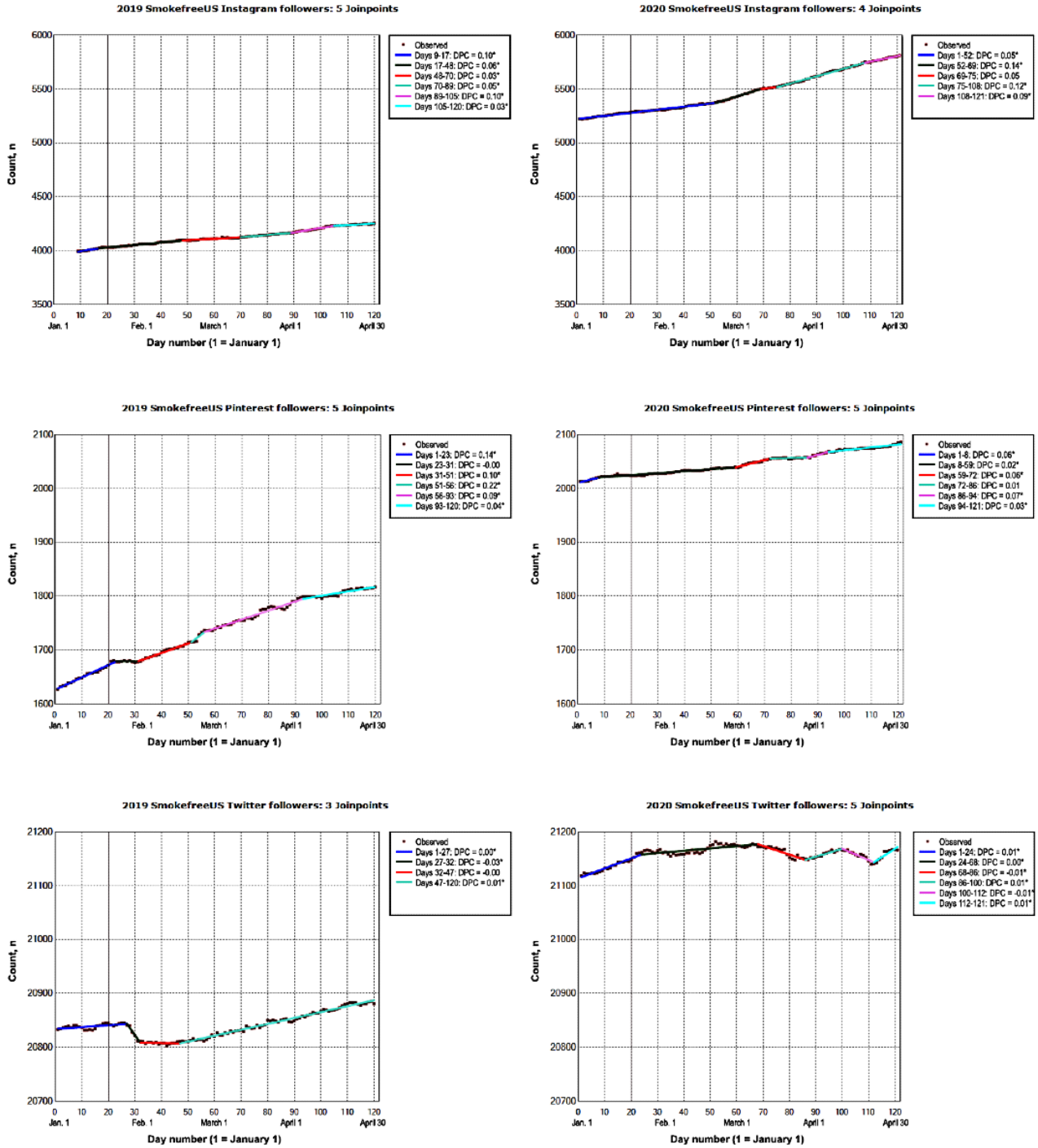


Figure 3. The number of SmokeFreeUS Instagram, Pinterest, and Twitter followers from January to April 2019 and 2020. The vertical line represents January 20. On this day in 2020, the first laboratory-confirmed COVID-19 case was identified in the United States. This was reported to the Centers for Disease Control and Prevention on January 22, 2020 [43]. Data on the number of daily SmokeFree Instagram followers were missing for the first 8 days in January 2019. Data on model selection appear in Multimedia Appendix 1. DPC: daily percent change. *The DPC is significantly different from 0 at an α level of .05.



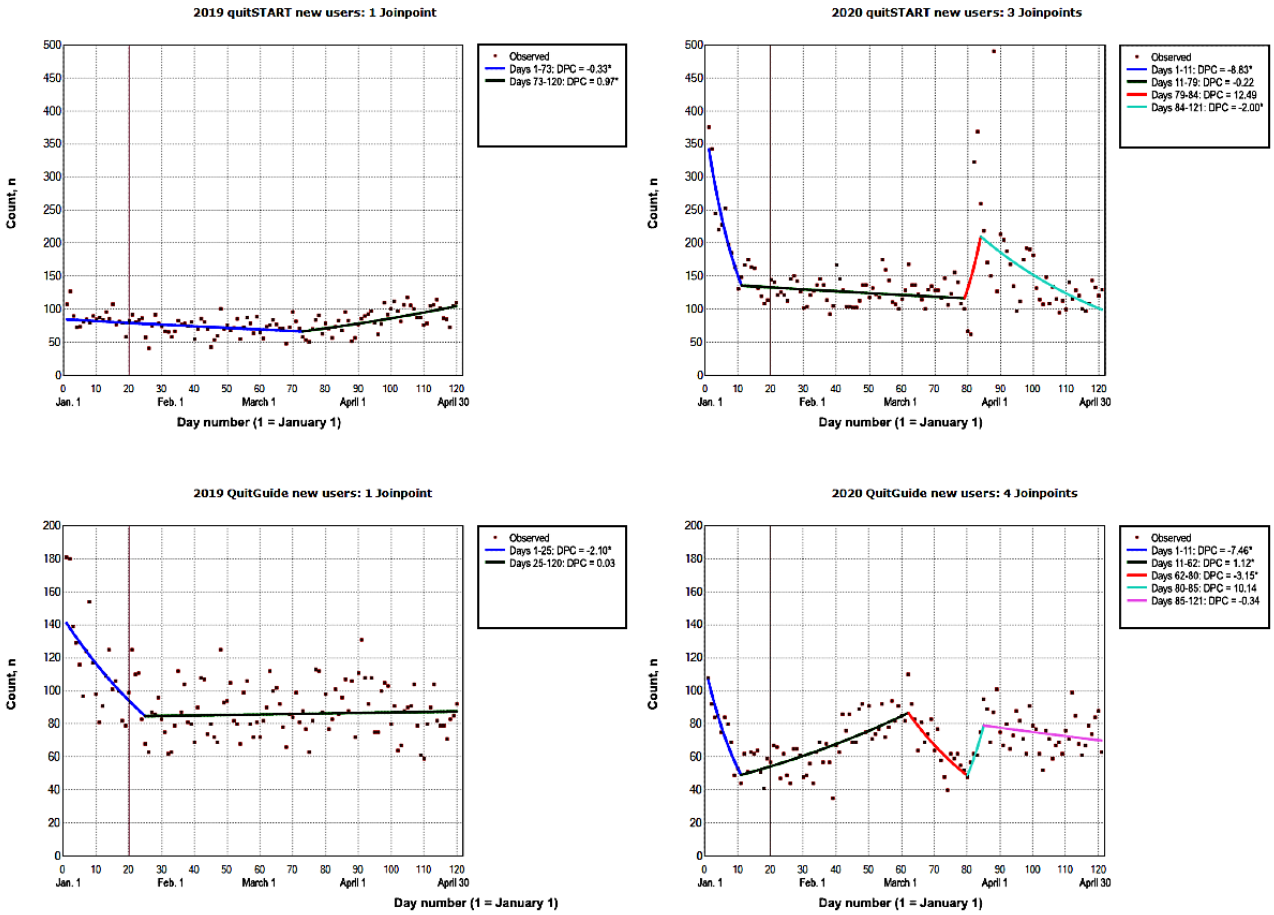
Smoking Cessation Interventions

Mobile Apps

In 2020, the number of new daily subscribers to quitSTART ranged from 62 (day 81) to 491 (day 88), and the number of new daily subscribers to QuitGuide ranged from 35 (day 39) to 110 (day 62; Figure 4). The number of new daily subscribers decreased across all intervention platforms until day 11 (quitSTART: DPC=−8.82%; 95% CI −12.82% to −4.64%; QuitGuide: DPC=−7.45%; 95% CI −10.64% to −4.15%; Table 1). The number of daily new subscribers to quitSTART did not significantly change from day 11 to day 84 (days 11-79: *P*=.08; days 79-84: *P*=.20). However, this number decreased by a DPC of −2.00% (95% CI −2.61% to −1.38%) from day 84 to day 121. For QuitGuide, the number of new daily subscribers increased by a DPC of 1.11% (95% CI 0.80% to 1.43%) from day 11 to day 62. Afterward, this number decreased by a DPC of −3.14% (95% CI −4.65% to −1.60%). From January to April

2020, the overall number of new daily subscribers to quitSTART trended downward (ADPC=−1.02%; 95% CI −1.88% to −0.15%), whereas the overall trend for QuitGuide (ADPC=−0.35%; 95% CI −1.06% to 0.37%) was not significant (*P*=.33; Table 2). In 2019, the number of new daily subscribers to quitSTART decreased by a DPC of −0.33% (95% CI −0.54% to −0.12%) between days 1 and 73, and the number of new subscribers to QuitGuide decreased by a DPC of −2.10% (95% CI −3.10% to −1.08) from day 1 to day 25. This was followed by a period of no significant changes in the number of daily subscribers to QuitGuide (*P*=.60). The overall number of new daily subscribers to QuitGuide trended downward from January to April 2019 (ADPC=−0.39%; 95% CI −0.62% to −0.17%), whereas the overall trend for quitSTART (ADPC=0.17%; 95% CI −0.01% to 0.37%) was not significant (*P*=.07).

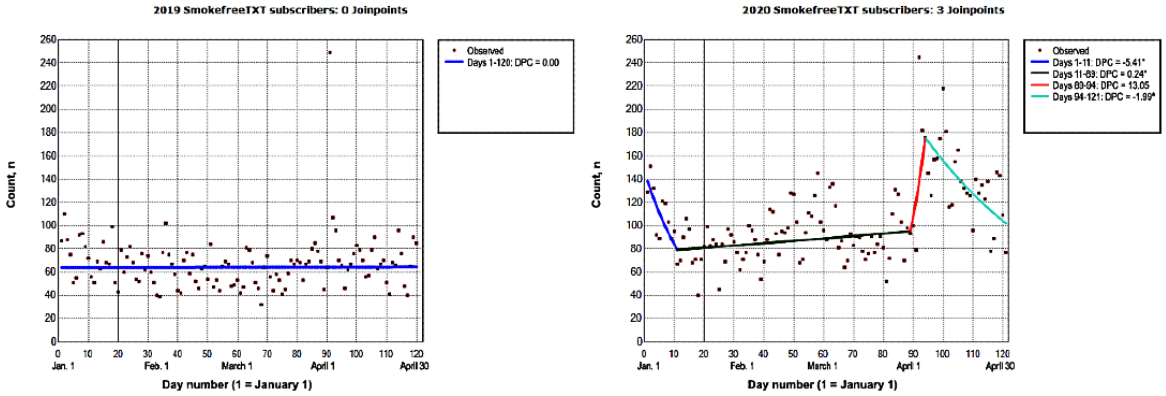
Figure 4. The number of new quitSTART and QuitGuide subscribers from January to April 2019 and 2020. The vertical line represents January 20. On this day in 2020, the first laboratory-confirmed COVID-19 case was identified in the United States. This was reported to the Centers for Disease Control and Prevention on January 22, 2020 [43]. Data on model selection appear in Multimedia Appendix 1. DPC: daily percent change. *The DPC is significantly different from 0 at an α level of .05.



Text Messaging

In 2020, the number of new daily subscribers to SmokeFreeTXT ranged from 40 (day 18) to 245 (day 92; Figure 5). The number of new daily subscribers to SmokeFreeTXT decreased by a DPC of −5.41% (95% CI −10.05% to −0.52%) until day 11 (Table 1). Additionally, the number of new daily subscribers to SmokeFreeTXT increased by a DPC of 0.23% (95% CI 0.004% to 0.47%) between days 11 and 89. This was followed by a period of no significant changes from day 89 to day 94 (*P*=.23) and a decrease in the number of new subscribers (DPC=−1.99%; 95% CI −3.09%, to −0.87%) from day 94 to day 121. From January to April 2020, the overall number of new daily subscribers to SmokeFreeTXT (ADPC=−0.25%; 95% CI −1.22% to 0.73%) was not significant (*P*=.61; Table 2). In 2019, the overall number of new daily subscribers to SmokeFreeTXT (ADPC=0.004%; 95% CI −0.14% to 0.15%) was not significant (*P*=.95).

Figure 5. The number of new SmokeFreeTXT subscribers from January to April 2019 and 2020. The vertical line represents January 20. On this day in 2020, the first laboratory-confirmed COVID-19 case was identified in the United States. This was reported to the Centers for Disease Control and Prevention on January 22, 2020 [43]. Data on model selection appear in Multimedia Appendix 1. DPC: daily percent change. *The DPC is significantly different from 0 at an α level of .05.



Discussion

Principal Findings

The COVID-19 pandemic is a natural event that has provided researchers with the opportunity to compare prepandemic and postpandemic digital platform traffic levels to understand the nature and magnitude of changes (if any) in the public's health perceptions and behaviors. This study showed that overall traffic trends on the SmokeFree website and adult-focused smoking cessation interventions fluctuated from January to April 2020. Although significant increases in the number of daily visitors and subscribers to these platforms were observed, they were eventually followed by declines in traffic. Contrastingly, there were no significant peaks in the 2019 traffic trends on the SmokeFree website and the adult-focused cessation interventions, suggesting that these trends were qualitatively different from those observed in 2020. The teen-focused smoking cessation intervention (ie, quitSTART) exhibited an overall downward trend in the number of new daily subscribers over the 4-month study period in 2020. The number of SmokeFree social media account followers modestly and steadily increased over the 4-month study periods in 2019 and 2020. Peaks in SmokeFree digital platform traffic in 2020 could be attributed to an increased interest in smoking cessation among smokers during the COVID-19 pandemic.

Digital platforms are beneficial for public health surveillance [44]. Previous research has shown that search queries, website traffic, and social media use reflect disease-related information-seeking behaviors, which can serve as timely indicators of infectious disease outbreaks (ie, during times when these behaviors deviate from normal patterns) [26,44-46]. In this study, the number of new daily visitors on the SmokeFree website increased significantly starting in early February 2020 (days 39-44: P=.006), and although traffic levels slightly decreased afterward, they remained higher than those of January. The trends observed in 2020 were qualitatively different from those observed in 2019; the number of new daily visitors on the website modestly increased throughout the 4-month study period in 2019. Elevated website traffic in 2020 can be viewed as a proxy indicator of interest in smoking cessation information, which is driven by several factors such as media coverage of the pandemic [47], smokers' perceived risk of SARS-CoV-2 infection [16], and calls from public health professionals to quit smoking as a preventive measure against SARS-CoV-2 infection [23,48]. Previously published literature has reported on dynamic changes in media coverage, information seeking behaviors, and perceived risk over the course of an epidemic and their importance in the adoption of precautionary behaviors and the acceptance of vaccinations [49,50].

The number of new daily subscriptions to SmokeFree cessation interventions generally peaked in early January. This is consistent with self-initiated behavioral changes that are triggered by naturally occurring events, such as New Year's Day [51,52]. After an initial decline from January levels, the number of new subscriptions to the adult-focused interventions QuitGuide and SmokeFreeTXT peaked and decreased throughout the remainder of the study period in 2020. This is consistent with previous research on people's precautionary behaviors, which faded over the course of the H1N1 epidemic [49]. Conversely, in 2019, the number of new subscriptions plateaued after an initial decline from January levels or remained steady throughout the study period. Although studies have suggested that the use of remote smoking cessation support services has increased during the COVID-19 pandemic [25], the fluctuations we observed in 2020 trends data could be attributed to several factors. First, evidence that links smoking to SARS-CoV-2 infection has evolved over time; several studies have suggested that nicotine is a protective/therapeutic factor against SARS-CoV-2 infection [53,54]. This prompted health organizations to provide cautionary statements [23]. Other studies have also provided evidence that smoking is a risk factor for SARS-CoV-2 infection [8-12]. Therefore, evolving evidence might have resulted in seemingly mixed or inconsistent public health messages at certain times. Second, lockdown measures and pandemic-associated economic hardships could have contributed to sustained smoking behaviors [55], which could have been reflected by the fluctuating cessation intervention uptake patterns that were observed in this study. Confinement-related psychological effects and stressors, which have been shown to increase in severity during previous epidemics and quarantines [56-58], are associated with smoking behaviors [59]. Indeed, research has shown that posttraumatic

depression and stress following natural disaster exposure (eg, hurricanes) indirectly result in smoking relapse [60]. With regard to COVID-19, studies have documented a desire to initiate tobacco use or relapse among nonsmokers and former smokers [61]. Similar self-reported increases in tobacco use have been documented in 30%-40% of e-cigarette users and cigar smokers during the COVID-19 pandemic [62,63].

The number of new daily subscriptions to the teen-focused SFGI intervention (ie, quitStart) exhibited an overall downward trend over the 2020 study period. Although there was a peak in the number of subscribers in mid- to late March, this peak was not statistically significant (days 79-84: P=.20). These results could be attributed to two factors. First, compared to adults, children aged <18 years are less susceptible to severe complications from COVID-19, including hospitalization and death [64-66]. This information could have resulted in the generation of reassuring reports (or lack thereof) on COVID-19 effects in youth. Reports of multisystem inflammatory syndrome in pediatric patients with COVID-19 emerged in March and April [67]. Any effects that these reports would have on traffic in digital smoking cessation platforms would have occurred beyond the January to April time frame of this study. Second, electronic vaping products are the most used products among high school students [68]. Accordingly, traffic on digital platforms for cigarette users might not be the best metric for analyzing interest in quit attempts among youth during the COVID-19 pandemic. Social media accounts can promote the web-based and non–web-based use of smoking cessation resources [69]. However, the number of SFGI social media followers was qualitatively similar across 2019 and 2020. Additional research is needed to understand the content of SFGI social media accounts during the pandemic. Furthermore, research is needed to understand people's web-based activities during the pandemic, including changes in social media following to induce behavioral change [70].

It should be noted that discrepancies in the timing and magnitude of increases and decreases in traffic on SmokeFree digital platforms could be attributed to several factors—mainly the medium of the platform, the aim of the public health effort, and the target audience. However, despite the modest increases and decreases in SmokeFree traffic, these changes are meaningful due to the reach of SmokeFree platforms [35]. It is well documented that the impact of public efforts (eg, interventions) are a product of reach and effectiveness [71]. Therefore, a mere 1% increase is meaningful if it represents hundreds or thousands of SmokeFree platform users. Furthermore, disseminating SmokeFree messages on multiple channels increases these messages' visibility and repitition, thereby increasing the likelihood of reaching a wide audience and providing higly effective messages that induce positive behavioral changes [72]. Based on findings of this study, coordinated smoking cessation campaigns, especially those for youth, should emphasize the importance of smoking cessation and raise awareness of digital smoking cessation platforms to capitalize on people's heightened interest during the pandemic. Public health professionals should also adapt pandemic-relevant messages and strategies instead of delivering generic smoking cessation interventions.

This study has several limitations. The platforms that were examined in this study are not representative of all digital smoking cessation resources (eg, BecomeAnEx) [73] or other resources that comply with physical distancing measures (eg, quitline) [74] and are available to smokers in the United States. Furthermore, the platforms examined in this study only target cigarette smokers, whereas resources available to other tobacco or nicotine users were not included (eg, text-to-quit vaping services) [75]. Future research should examine traffic on alternative tobacco or nicotine product cessation platforms (eg, e-cigarette use among teens and young adults) and surveil nontraditional sources (eg, pharmaceutical sales of smoking cessation aids and tobacco products sales). Such research will complement our data and provide a comprehensive picture of tobacco use during the COVID-19 pandemic [76]. Additionally, website traffic, social media followers, and intervention subscriptions are not indicative of successful quit attempts. Future research should examine smoking abstinence, reductions in the number of cigarettes smoked, and product switching (eg, switching from cigarettes to e-cigarettes) among users of SFGI digital platforms during the pandemic, as evidence on people's motivation to quit and quitting success rates during the pandemic is mixed [62,77,78]. Traffic on SFGI platforms reflects the number of smokers who are self-motivated to quit and have access to computers/mobile phones with broadband, data, and text messaging plans. Therefore, plans for waiving smokers' fees or providing data/text messaging plans to smokers during their quit journey should be implemented, so that smokers who might lack resources can access smoking cessation digital platforms. Furthermore, research is needed to understand the racial/ethnic and socioeconomic profiles of SFGI users. Such data were not available for analysis in this study (eg, there is evidence of increased COVID-19 exposure rates and severity risk among Black and Hispanic people) [79]. Of note, several factors that could affect traffic on SmokeFree digital platforms were not considered in this study (eg, activities for promoting SmokeFree platforms, the volume of media reports on COVID-19 and smoking, etc). This study examined prepandemic and postpandemic traffic on SmokeFree platforms over 4 months. Future research should extend the time and geographic boundaries of our study to identify smoking cessation digital platform traffic trends that occur beyond these initial 4 months in countries other than the United States.

Conclusion

This study characterized traffic trends on SFGI digital platforms from January to April 2019 and 2020. Traffic on the SmokeFree website and adult-focused interventions increased in mid-January and February 2020, whereas traffic on the teen-focused intervention exhibited an overall downward trend. Comparable trends were not observed in 2019. The number of social media followers was similar across the 2019 and 2020 study periods. The 2020 traffic trends on the SFGI website and intervention platforms reflected opposing dynamics in the relationship between people's interest in smoking cessation as a preventive measure against COVID-19 and evolving evidence on the risk profiles of patients with COVID-19 who develop adverse outcomes, the link between smoking and COVID-19–related complications, and sustained/increased smoking due to pandemic-related stressors.

Acknowledgments

Analyses were conducted by Information Management Services Inc, which is an information management services firm with a specialty in statistical consultation. This firm is used by the Division of Intramural Research, National Institute on Minority Health and Health Disparities. The effort of SET was supported by the Division of Intramural Research, National Institute on Minority Health and Health Disparities, National Institutes of Health. The SmokeFree website is funded by the Tobacco Control Research Branch of the National Cancer Institute, National Institutes of Health (contract numbers: HHSN261201400002B and HHSN261000011).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Results of the joinpoint regression analysis.
[DOCX File , 15 KB - jmir_v23i3e24593_app1.docx]

References

1. Rolling updates on coronavirus disease (covid-19). World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen [accessed 2020-06-04]
2. Johns Hopkins coronavirus resource center. Johns Hopkins University & Medicine. URL: https://coronavirus.jhu.edu/ [accessed 2020-06-04]
3. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
4. U.S. Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta, GA: Centers for Disease Control and Prevention (US); 2014.
5. Zhou Z, Chen P, Peng H. Are healthy smokers really healthy? Tob Induc Dis 2016 Nov 15;14:35 [FREE Full text] [doi: 10.1186/s12971-016-0101-z] [Medline: 27891067]
6. Arcavi L, Benowitz NL. Cigarette smoking and infection. Arch Intern Med 2004 Nov 08;164(20):2206-2216. [doi: 10.1001/archinte.164.20.2206] [Medline: 15534156]
7. Park JE, Jung S, Kim A, Park JE. MERS transmission and risk factors: a systematic review. BMC Public Health 2018 May 02;18(1):574 [FREE Full text] [doi: 10.1186/s12889-018-5484-8] [Medline: 29716568]
8. Patanavanich R, Glantz SA. Smoking is associated with COVID-19 progression: A meta-analysis. Nicotine Tob Res 2020 Aug 24;22(9):1653-1656 [FREE Full text] [doi: 10.1093/ntr/ntaa082] [Medline: 32399563]
9. Vardavas CI, Nikitara K. COVID-19 and smoking: A systematic review of the evidence. Tob Induc Dis 2020 Mar 20;18:20 [FREE Full text] [doi: 10.18332/tid/119324] [Medline: 32206052]
10. Zhao Q, Meng M, Kumar R, Wu Y, Huang J, Lian N, et al. The impact of COPD and smoking history on the severity of COVID-19: A systemic review and meta-analysis. J Med Virol 2020 Oct;92(10):1915-1921 [FREE Full text] [doi: 10.1002/jmv.25889] [Medline: 32293753]
11. Alqahtani JS, Oyelade T, Aldhahir AM, Alghamdi SM, Almehmadi M, Alqahtani AS, et al. Prevalence, severity and mortality associated with COPD and smoking in patients with COVID-19: A rapid systematic review and meta-analysis. PLoS One 2020 May 11;15(5):e0233147. [doi: 10.1371/journal.pone.0233147] [Medline: 32392262]
12. Emami A, Javanmardi F, Pirbonyeh N, Akbari A. Prevalence of Underlying Diseases in Hospitalized Patients with COVID-19: a Systematic Review and Meta-Analysis. Arch Acad Emerg Med 2020;8(1):e35 [FREE Full text] [Medline: 32232218]
13. Simons D, Shahab L, Brown J, Perski O. The association of smoking status with SARS-CoV-2 infection, hospitalization and mortality from COVID-19: a living rapid evidence review with Bayesian meta-analyses (version 7). Addiction. 2020 Oct 02 Epub ahead of print. [FREE Full text] [doi: 10.1111/add.15276] [Medline: 33007104]
14. Lippi G, Henry BM. Active smoking is not associated with severity of coronavirus disease 2019 (COVID-19). Eur J Intern Med 2020 May;75:107-108 [FREE Full text] [doi: 10.1016/j.ejim.2020.03.014] [Medline: 32192856]
15. Komiyama M, Hasegawa K. Smoking cessation as a public health measure to limit the coronavirus disease 2019 pandemic. Eur Cardiol 2020 Apr 23;15:e16 [FREE Full text] [doi: 10.15420/ecr.2020.11] [Medline: 32373189]
16. Jackson SE, Brown J, Shahab L, Steptoe A, Fancourt D. COVID-19, smoking,inequalities: a cross-sectional survey of adults in the UK. medRxiv. Preprint posted online on May 5, 2020 [FREE Full text] [doi: 10.1101/2020.04.30.20086074]
17. Sheeran P, Harris PR, Epton T. Does heightening risk appraisals change people's intentions and behavior? A meta-analysis of experimental studies. Psychol Bull 2014 Mar;140(2):511-543. [doi: 10.1037/a0033065] [Medline: 23731175]
18. McBride CM, Emmons KM, Lipkus IM. Understanding the potential of teachable moments: the case of smoking cessation. Health Educ Res 2003 Apr;18(2):156-170. [doi: 10.1093/her/18.2.156] [Medline: 12729175]

19. Tonnesen P, Marott JL, Nordestgaard B, Bojesen SE, Lange P. Secular trends in smoking in relation to prevalent and incident smoking-related disease: A prospective population-based study. Tob Induc Dis 2019 Oct 07;17:72 [FREE Full text] [doi: 10.18332/tid/112459] [Medline: 31768164]
20. Danielsen SE, Løchen ML, Medbø A, Vold ML, Melbye H. A new diagnosis of asthma or COPD is linked to smoking cessation - the Tromsø study. Int J Chron Obstruct Pulmon Dis 2016 Jun 30;11:1453-1458. [doi: 10.2147/COPD.S108046] [Medline: 27418818]
21. Berlin I, Thomas D, Le Faou AL, Cornuz J. COVID-19 and smoking. Nicotine Tob Res 2020 Aug 24;22(9):1650-1652 [FREE Full text] [doi: 10.1093/ntr/ntaa059] [Medline: 32242236]
22. Eisenberg SL, Eisenberg MJ. Smoking cessation during the COVID-19 epidemic. Nicotine Tob Res 2020 Aug 24;22(9):1664-1665 [FREE Full text] [doi: 10.1093/ntr/ntaa075] [Medline: 32363386]
23. WHO statement: Tobacco use and covid-19. World Health Organization. URL: https://www.who.int/news-room/detail/11-05-2020-who-statement-tobacco-use-and-covid-19 [accessed 2020-06-25]
24. Kapoor A, Guha S, Das MK, Goswami KC, Yadav R. Digital healthcare: The only solution for better healthcare during COVID-19 pandemic? Indian Heart J 2020;72(2):61-64 [FREE Full text] [doi: 10.1016/j.ihj.2020.04.001] [Medline: 32534691]
25. Jackson SE, Garnett C, Shahab L, Oldham M, Brown J. Association of the Covid-19 lockdown with smoking, drinking, and attempts to quit in England: An analysis of 2019-2020 data. medRxiv. Preprint posted online on May 27, 2020 [FREE Full text] [doi: 10.1101/2020.05.25.20112656]
26. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: A systematic review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
27. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014;9(10):e109583. [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
28. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev. 2019 Oct 22;10(10) Epub ahead of print. [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
29. Heerfordt C, Heerfordt IM. Has there been an increased interest in smoking cessation during the first months of the COVID-19 pandemic? A Google Trends study. Public Health 2020 Jun;183:6-7 [FREE Full text] [doi: 10.1016/j.puhe.2020.04.012] [Medline: 32388011]
30. #quit4covid. Twitter. URL: https://twitter.com/hashtag/quit4covid?lang=en [accessed 2020-06-10]
31. Kuchler T, Russel D, Stroebel J. The geographic spread of COVID-19 correlates with structure of social networks as measured by Facebook. National Bureau of Economic Research. 2020. URL: https://www.nber.org/system/files/working_papers/w26990/w26990.pdf [accessed 2021-03-04]
32. Perski O, Herbeć A, Shahab L, Brown J. Influence of the SARS-CoV-2 outbreak on the uptake of a popular smoking cessation app in UK smokers: Interrupted time series analysis. JMIR Mhealth Uhealth 2020 Jun 11;8(6):e19494 [FREE Full text] [doi: 10.2196/19494] [Medline: 32463375]
33. Smokefree. National Cancer Institute. URL: https://smokefree.gov/ [accessed 2020-06-11]
34. Businelle MS. The potential of mHealth for tobacco dependence treatment: domestic and international examples from NCI's Smokefree.gov initiative. Nicotine Tob Res 2014 Jul;16(7):1033. [doi: 10.1093/ntr/ntu071] [Medline: 24838843]
35. Prutzman YM, Wiseman KP, Grady MA, Budenz A, Grenen EG, Vercammen LK, et al. Using digital technologies to reach tobacco users who want to quit: Evidence from the National Cancer Institute's Smokefree.gov Initiative. Am J Prev Med 2021 Mar;60(3 Suppl 2):S172-S184 [FREE Full text] [doi: 10.1016/j.amepre.2020.08.008] [Medline: 33663705]
36. How to quit. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/tobacco/quit_smoking/how_to_quit/index.htm [accessed 2020-08-26]
37. Quitting smoking: Closer with every attempt. U.S. Food & Drug Administration. URL: https://www.fda.gov/tobacco-products/health-information/quitting-smoking-closer-every-attempt [accessed 2020-08-26]
38. Michie S, Hyder N, Walia A, West R. Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. Addict Behav 2011 Apr;36(4):315-319. [doi: 10.1016/j.addbeh.2010.11.016] [Medline: 21215528]
39. Augustson E, Cole-Lewis H, Sanders A, Schwarz M, Geng Y, Coa K, et al. Analysing user-reported data for enhancement of SmokefreeTXT: a national text message smoking cessation intervention. Tob Control 2017 Nov;26(6):683-689. [doi: 10.1136/tobaccocontrol-2016-052945] [Medline: 27852892]
40. Jointpoint regression program. National Cancer Institute. URL: https://surveillance.cancer.gov/joinpoint/ [accessed 2021-03-04]
41. Ingram DD, Malec DJ, Makuc DM, Kruszon-Moran D, Gindi RM, Albert M, et al. National Center for Health Statistics Guidelines for Analysis of Trends. Vital Health Stat 2 2018 Apr(179):1-71 [FREE Full text] [Medline: 29775435]
42. Number of joinpoints. National Cancer Institute. URL: https://surveillance.cancer.gov/help/joinpoint/setting-parameters/method-and-parameters-tab/number-of-joinpoints [accessed 2021-03-17]

43. Stokes EK, Zambrano LD, Anderson KN, Marder EP, Raz KM, Felix SEB, et al. Coronavirus disease 2019 case surveillance - United States, January 22-May 30, 2020. MMWR Morb Mortal Wkly Rep 2020 Jun 19;69(24):759-765 [FREE Full text] [doi: 10.15585/mmwr.mm6924e2] [Medline: 32555134]
44. Barros JM, Duggan J, Rebholz-Schuhmann D. The application of internet-based sources for public health surveillance (infoveillance): Systematic review. J Med Internet Res 2020 Mar 13;22(3):e13680 [FREE Full text] [doi: 10.2196/13680] [Medline: 32167477]
45. Gittelman S, Lange V, Crawford CAG, Okoro CA, Lieb E, Dhingra SS, et al. A new source of data for public health surveillance: Facebook likes. J Med Internet Res 2015 Apr 20;17(4):e98 [FREE Full text] [doi: 10.2196/jmir.3970] [Medline: 25895907]
46. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006;2006:244-248 [FREE Full text] [Medline: 17238340]
47. Sousa-Pinto B, Anto A, Czarlewski W, Anto JM, Fonseca JA, Bousquet J. Assessment of the impact of media coverage on COVID-19-related google trends data: Infodemiology study. J Med Internet Res 2020 Aug 10;22(8):e19611 [FREE Full text] [doi: 10.2196/19611] [Medline: 32530816]
48. #HealthAtHome - Quitting tobacco. World Health Organization. URL: https://www.who.int/campaigns/connecting-the-world-to-combat-coronavirus/healthyathome/healthyathome---quitting-tobacco [accessed 2020-08-28]
49. Ibuka Y, Chapman GB, Meyers LA, Li M, Galvani AP. The dynamics of risk perceptions and precautionary behavior in response to 2009 (H1N1) pandemic influenza. BMC Infect Dis 2010 Oct 14;10:296 [FREE Full text] [doi: 10.1186/1471-2334-10-296] [Medline: 20946662]
50. Walter D, Bohmer M, Reiter S, Krause G, Wichmann O. Risk perception and information-seeking behaviour during the 2009/10 influenza A(H1N1)pdm09 pandemic in Germany. Euro Surveill 2012 Mar 29;17(13):20131 [FREE Full text] [Medline: 22490383]
51. Norcross JC, Ratzin AC, Payne D. Ringing in the new year: the change processes and reported outcomes of resolutions. Addict Behav 1989;14(2):205-212. [doi: 10.1016/0306-4603(89)90050-6] [Medline: 2728957]
52. Graham AL, Cha S, Cobb NK, Fang Y, Niaura RS, Mushro A. Impact of seasonality on recruitment, retention, adherence, and outcomes in a web-based smoking cessation intervention: randomized controlled trial. J Med Internet Res 2013 Nov 07;15(11):e249 [FREE Full text] [doi: 10.2196/jmir.2880] [Medline: 24201304]
53. Farsalinos K, Barbouni A, Niaura R. Systematic review of the prevalence of current smoking among hospitalized COVID-19 patients in China: could nicotine be a therapeutic option? Intern Emerg Med 2020 Aug;15(5):845-852 [FREE Full text] [doi: 10.1007/s11739-020-02355-7] [Medline: 32385628]
54. Tindle HA, Newhouse PA, Freiberg MS. Beyond smoking cessation: Investigating medicinal nicotine to prevent and treat COVID-19. Nicotine Tob Res 2020 Aug 24;22(9):1669-1670 [FREE Full text] [doi: 10.1093/ntr/ntaa077] [Medline: 32383751]
55. Patwardhan P. COVID-19: Risk of increase in smoking rates among England's 6 million smokers and relapse among England's 11 million ex-smokers. BJGP Open 2020 Jun 23;4(2):1-3 [FREE Full text] [doi: 10.3399/bjgpopen20X101067] [Medline: 32265183]
56. Hawryluck L, Gold WL, Robinson S, Pogorski S, Galea S, Styra R. SARS control and psychological effects of quarantine, Toronto, Canada. Emerg Infect Dis 2004 Jul;10(7):1206-1212 [FREE Full text] [doi: 10.3201/eid1007.030703] [Medline: 15324539]
57. Jeong H, Yim HW, Song YJ, Ki M, Min JA, Cho J, et al. Mental health status of people isolated due to Middle East Respiratory Syndrome. Epidemiol Health 2016 Nov 05;38:e2016048. [doi: 10.4178/epih.e2016048] [Medline: 28196409]
58. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
59. Perkins KA, Grobe JE. Increased desire to smoke during acute stress. Br J Addict 1992 Jul;87(7):1037-1040. [doi: 10.1111/j.1360-0443.1992.tb03121.x] [Medline: 1643396]
60. Alexander AC, Ward KD, Forde DR, Stockton M. Are posttraumatic stress and depressive symptoms pathways to smoking relapse after a natural disaster? Drug Alcohol Depend 2019 Feb 01;195:178-185. [doi: 10.1016/j.drugalcdep.2018.09.025] [Medline: 30455073]
61. Caponnetto P, Inguscio L, Saitta C, Maglia M, Benfatto F, Polosa R. Smoking behavior and psychological dynamics during COVID-19 social distancing and stay-at-home policies: A survey. Health Psychol Res 2020 May 27;8(1):9124 [FREE Full text] [doi: 10.4081/hpr.2020.9124] [Medline: 32529094]
62. Klemperer EM, West JC, Peasley-Miklus C, Villanti AC. Change in tobacco and electronic cigarette use and motivation to quit in response to COVID-19. Nicotine Tob Res 2020 Aug 24;22(9):1662-1663 [FREE Full text] [doi: 10.1093/ntr/ntaa072] [Medline: 32343816]
63. Kowitt SD, Ross JC, Jarman KL, Kistler CE, Lazard AJ, Ranney LM, et al. Tobacco quit intentions and behaviors among cigar smokers in the United States in response to COVID-19. Int J Environ Res Public Health 2020 Jul 25;17(15):5368 [FREE Full text] [doi: 10.3390/ijerph17155368] [Medline: 32722469]

64. Dong Y, Mo X, Hu Y, Qi X, Jiang F, Jiang Z, et al. Epidemiology of COVID-19 among children in China. Pediatrics 2020 Jun;145(6):e20200702. [doi: 10.1542/peds.2020-0702] [Medline: 32179660]
65. CDC COVID-19 Response Team. Coronavirus disease 2019 in children - United States, February 12-April 2, 2020. MMWR Morb Mortal Wkly Rep 2020 Apr 10;69(14):422-426 [FREE Full text] [doi: 10.15585/mmwr.mm6914e4] [Medline: 32271728]
66. Götzinger F, Santiago-García B, Noguera-Julián A, Lanaspa M, Lancella L, Carducci FIC, ptbnet COVID-19 Study Group. COVID-19 in children and adolescents in Europe: a multinational, multicentre cohort study. Lancet Child Adolesc Health 2020 Sep;4(9):653-661 [FREE Full text] [doi: 10.1016/S2352-4642(20)30177-2] [Medline: 32593339]
67. Godfred-Cato S, Bryant B, Leung J, Oster ME, Conklin L, Abrams J, California MIS-C Response Team. COVID-19-associated multisystem inflammatory syndrome in children - United States, March-July 2020. MMWR Morb Mortal Wkly Rep 2020 Aug 14;69(32):1074-1080 [FREE Full text] [doi: 10.15585/mmwr.mm6932e2] [Medline: 32790663]
68. Creamer MR, Jones SE, Gentzke AS, Jamal A, King BA. Tobacco product use among high school students - Youth risk behavior survey, United States, 2019. MMWR Suppl 2020 Aug 21;69(1):56-63 [FREE Full text] [doi: 10.15585/mmwr.su6901a7] [Medline: 32817607]
69. Duke JC, Hansen H, Kim AE, Curry L, Allen J. The use of social media by state tobacco control programs to promote smoking cessation: a cross-sectional study. J Med Internet Res 2014 Jul 10;16(7):e169 [FREE Full text] [doi: 10.2196/jmir.3430] [Medline: 25014311]
70. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
71. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
72. Making health communication programs work: A planner's guide. National Cancer Institute. URL: https://www.cancer.gov/publications/health-communication/pink-book.pdf [accessed 2021-02-08]
73. Become an EX smoker, learn to quit smoking, stop smoking cigarettes. BecomeAnEX. URL: https://www.becomeanex.org/ [accessed 2020-08-24]
74. NAQC quitline map. North America Quitline Consortium Quitline Map. URL: http://map.naquitline.org/ [accessed 2020-08-24]
75. This is quitting. Truth. URL: https://www.thetruth.com/about-truth/collaboration/this-is-quitting [accessed 2020-08-24]
76. Metzger KB, Mostashari F, Kerker BD. Use of pharmacy data to evaluate smoking regulations' impact on sales of nicotine replacement therapies in New York City. Am J Public Health 2005 Jun;95(6):1050-1055 [FREE Full text] [doi: 10.2105/AJPH.2004.048025] [Medline: 15914832]
77. Tetik BK, Tekinemre IG, Taş S. The effect of the COVID-19 pandemic on smoking cessation success. J Community Health. 2020 Jul 08 Epub ahead of print. [FREE Full text] [doi: 10.1007/s10900-020-00880-2] [Medline: 32643078]
78. Elling JM, Crutzen R, Talhout R, de Vries H. Tobacco smoking and smoking cessation in times of COVID-19. Tob Prev Cessat 2020 Jul 01;6:39 [FREE Full text] [doi: 10.18332/tpc/122753] [Medline: 33083672]
79. Hooper MW, Nápoles AM, Pérez-Stable EJ. COVID-19 and racial/ethnic disparities. JAMA 2020 Jun 23;323(24):2466-2467. [doi: 10.1001/jama.2020.8598] [Medline: 32391864]

Abbreviations

ADPC: average daily percentage change
DPC: daily percent change
SFGI: SmokeFree.gov initiative

Edited by G Eysenbach, C Basch; submitted 25.09.20; peer-reviewed by O Perski, J Chan, D Cheung; comments to author 30.11.20; revised version received 31.12.20; accepted 19.02.21; published 22.03.21.

Please cite as:
El-Toukhy S
Insights From the SmokeFree.gov Initiative Regarding the Use of Smoking Cessation Digital Platforms During the COVID-19 Pandemic: Cross-sectional Trends Analysis Study
J Med Internet Res 2021;23(3):e24593
URL: https://www.jmir.org/2021/3/e24593
doi:10.2196/24593
PMID:33646963

Original Paper

Short-Range Forecasting of COVID-19 During Early Onset at County, Health District, and State Geographic Levels Using Seven Methods: Comparative Forecasting Study

Christopher J Lynch[1], PhD; Ross Gore[1], PhD

Virginia Modeling, Analysis, and Simulation Center, Old Dominion University, Suffolk, VA, United States

Corresponding Author:
Christopher J Lynch, PhD
Virginia Modeling, Analysis, and Simulation Center
Old Dominion University
1030 University Blvd
Suffolk, VA, 23435
United States
Phone: 1 7576866248
Email: cjlynch@odu.edu

Abstract

Background: Forecasting methods rely on trends and averages of prior observations to forecast COVID-19 case counts. COVID-19 forecasts have received much media attention, and numerous platforms have been created to inform the public. However, forecasting effectiveness varies by geographic scope and is affected by changing assumptions in behaviors and preventative measures in response to the pandemic. Due to time requirements for developing a COVID-19 vaccine, evidence is needed to inform short-term forecasting method selection at county, health district, and state levels.

Objective: COVID-19 forecasts keep the public informed and contribute to public policy. As such, proper understanding of forecasting purposes and outcomes is needed to advance knowledge of health statistics for policy makers and the public. Using publicly available real-time data provided online, we aimed to evaluate the performance of seven forecasting methods utilized to forecast cumulative COVID-19 case counts. Forecasts were evaluated based on how well they forecast 1, 3, and 7 days forward when utilizing 1-, 3-, 7-, or all prior–day cumulative case counts during early virus onset. This study provides an objective evaluation of the forecasting methods to identify forecasting model assumptions that contribute to lower error in forecasting COVID-19 cumulative case growth. This information benefits professionals, decision makers, and the public relying on the data provided by short-term case count estimates at varied geographic levels.

Methods: We created 1-, 3-, and 7-day forecasts at the county, health district, and state levels using (1) a naïve approach, (2) Holt-Winters (HW) exponential smoothing, (3) a growth rate approach, (4) a moving average (MA) approach, (5) an autoregressive (AR) approach, (6) an autoregressive moving average (ARMA) approach, and (7) an autoregressive integrated moving average (ARIMA) approach. Forecasts relied on Virginia's 3464 historical county-level cumulative case counts from March 7 to April 22, 2020, as reported by *The New York Times*. Statistically significant results were identified using 95% CIs of median absolute error (MdAE) and median absolute percentage error (MdAPE) metrics of the resulting 216,698 forecasts.

Results: The next-day MA forecast with 3-day look-back length obtained the lowest MdAE (median 0.67, 95% CI 0.49-0.84, $P<.001$) and statistically significantly differed from 39 out of 59 alternatives (66%) to 53 out of 59 alternatives (90%) at each geographic level at a significance level of .01. For short-range forecasting, methods assuming stationary means of prior days' counts outperformed methods with assumptions of weak stationarity or nonstationarity means. MdAPE results revealed statistically significant differences across geographic levels.

Conclusions: For short-range COVID-19 cumulative case count forecasting at the county, health district, and state levels during early onset, the following were found: (1) the MA method was effective for forecasting 1-, 3-, and 7-day cumulative case counts; (2) exponential growth was not the best representation of case growth during early virus onset when the public was aware of the virus; and (3) geographic resolution was a factor in the selection of forecasting methods.

(J Med Internet Res 2021;23(3):e24925) doi:10.2196/24925

KEYWORDS

coronavirus disease 2019; COVID-19; infectious disease; emerging outbreak; forecasting; modeling and simulation; public health; modeling disease outbreaks

Introduction

The scientific community responded quickly to the global outbreak following COVID-19's identification in December of 2019 [1,2]. Numerous platforms and studies have been created to forecast the spread of the pandemic and meet the need for intervention measures in support of public health and awareness [1,3-5]. Many forecasting efforts focused on the long-term identification of COVID-19 and the flattening of and getting over the curve [3,4,6]. Forecasts assist in identifying and evaluating long-term preventions. Short-range forecasts provide benefit by supporting local understanding for individuals and policy makers and supporting short-range decisions. To inform public health and support awareness of proper forecast interpretation when making decisions, it is important to understand the basics of the generation of forecasts and the boundaries under which their interpretations are valid. To this point, this study explores the error levels of seven common forecasting methods in estimating COVID-19 cumulative case counts at county, health district, and state levels over the upcoming week. Comparing error levels across forecasting methods and geographic granularities provides insight into the assumptions contributing to more accurate forecasts.

Small numbers of COVID-19 cases can lead to large outbreaks [7]. Isolation and preventative measures are recommended practices to reduce the spread of COVID-19 [1-3,8-14]. Forecasts with high error magnitudes can provide expectations that grossly underestimate or overestimate case counts. This can lead to problems, such as the creation of unanticipated hot spots resulting from underestimation, or can cause unnecessary public alarm from overestimation. Interpreting COVID-19 forecasts depends on assumptions such as the geographic area, preventative measures in place, and the population's knowledge of, and behaviors toward, the virus. As assumptions change, the usefulness of the forecasting method should be re-evaluated. The impact of nonpharmaceutical interventions can be delayed 1 to 3 weeks and should factor into policy makers' decisions [15]. Intervention methods can result in secondary effects, such as decreasing levels of physical activity while people practice social distancing [16]. As a result, understanding of the assumptions pertaining to short-range COVID-19 forecasting is needed to properly interpret their findings [17].

This study explores seven commonly utilized forecasting approaches, including the following: naïve [18], moving average (MA) [9,10], autoregressive (AR) [17], growth rate [19], Holt-Winters (HW) exponential smoothing [20,21], autoregressive moving average (ARMA) [22], and autoregressive integrated moving average (ARIMA) [23]. Each forecasting method utilizes different assumptions about how the past values impact the forecast values. The naïve approach is the simplest method and assumes no change from the current value. The MA approach assumes equal weighting of prior values, while exponential smoothing assigns exponentially decreasing weight to older values. The AR approach assumes linear dependency of prior values but with an added stochastic component. The growth rate approach assumes a linear relationship to its prior values and applies sampling with growth based on the number of increased cases from the prior day. The ARMA approach combines the AR approach to provide a regression based linearly on its past values with the MA approach to account for the error terms within the prior values. The ARIMA approach applies to data that is nonstationary around a mean value and applies a distancing measure one or more times to make the data stationary [24].

Error represents the inability to account for all the variability contributing to changes in COVID-19 case counts. Forecast error represents the under- or overestimation of the actual value [18]. Additionally, assumptions are unlikely to remain constant over time due to shifting public behaviors and implemented public policies. Error magnitude communicates the accuracy of a forecast and can be utilized as a metric to select from a set of potential forecasting methods. Interpreting forecast outcomes relies on the error magnitude as well as situating the assumptions underlining the forecast [25,26]. This means that the effectiveness of a current forecasting method is likely to be impacted as new preventative measures are put into place that alter spread dynamics. Conveying this understanding to the public advances knowledge of health statistics and statistical literacy in public health [21,27-29].

Recommendations for models of infectious diseases in support of public health involve incorporating policy questions, available data, and scientific understanding to yield policy advice, data collection, and scientific insight [30]. By evaluating forecasts by aggregating information from lower levels, such as at the county and health district levels, intervention strategies can be more readily applied based on the relevant demographic characteristics shared by the smaller population samples. Forecasting methods operate under differing assumptions pertaining to how the prior values relate to forecasted values. This study evaluates seven forecasting methods with varied look-back and forecast lengths at the county, health district, and state levels. By evaluating short-range forecasting methods combined with varied look-back and forecast lengths, forecasts can be more effectively used for informing health planning and aiding individuals in evaluating the safety levels of their local and neighboring communities.

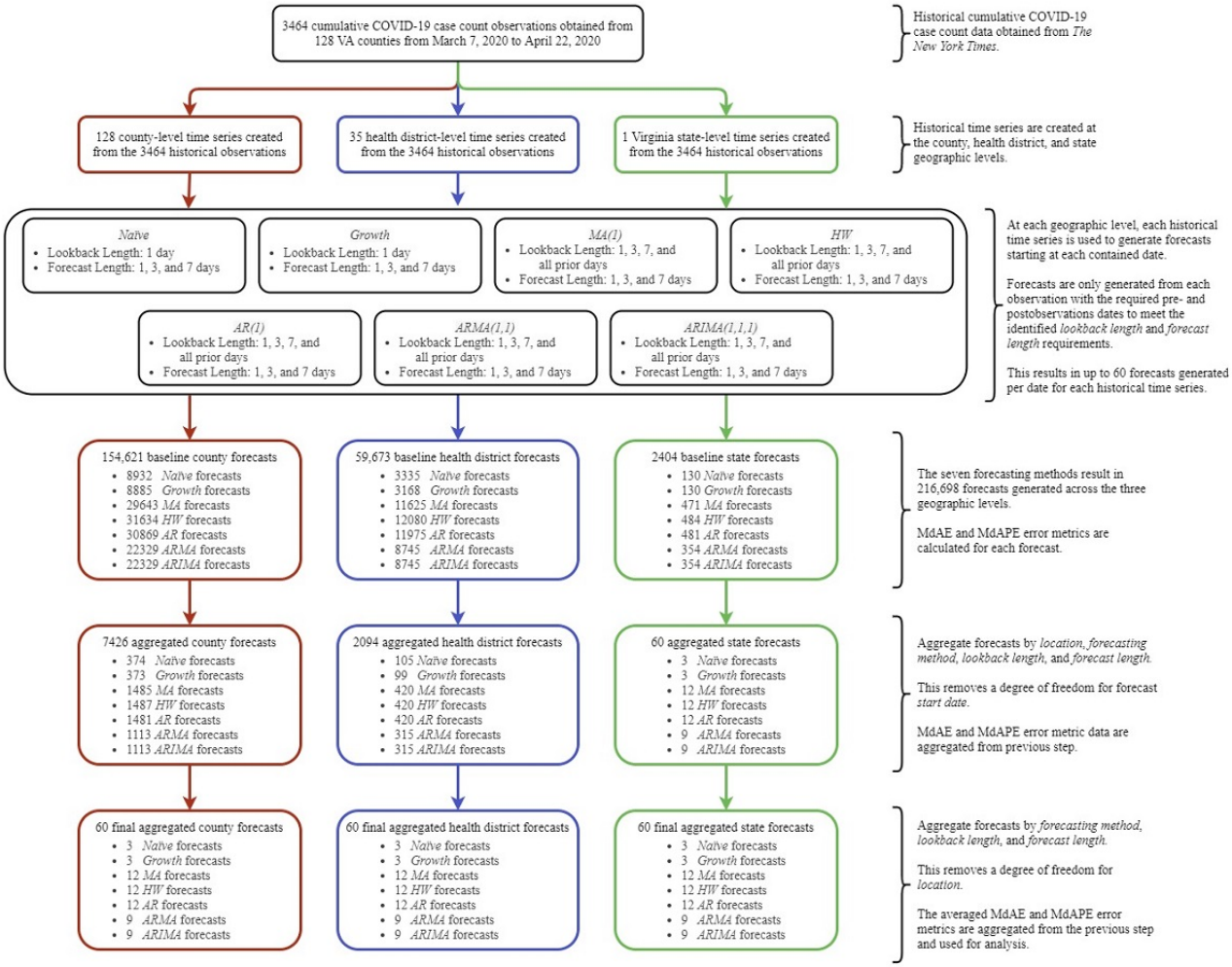
Methods

Data

We obtained 3464 Virginia county–level COVID-19 cumulative case count observations from March 7 to April 22, 2020, using data provided by *The New York Times* and aggregated these observations to the health district and state levels as presented in Figure 1. This period captured the first 3 weeks following the first confirmed COVID-19 case within Virginia and the 3 weeks following the governor's executive order limiting gatherings to groups of less than 10 people. As intervention

measures can take up to 3 weeks to impact the virus spread [15], this time frame was expected to cover Virginia's case growth prior to experiencing the benefits resulting from the governor's imposed group size limit.

Figure 1. Experimental design and data overview at the county, health district, and state levels. The generation and aggregation of county-level forecasts are shown on the left path (red), health district–level forecasts on the middle path (blue), and state-level forecasts on the right path (green). The information on the right provides additional detail on each stage in the experimental design. AR: autoregressive; ARIMA: autoregressive integrated moving average; ARMA: autoregressive moving average; HW: Holt-Winters; MA: moving average; MdAE: median absolute error; MdAPE: median absolute percentage error; VA: Virginia.



Forecasting Methods and Assumptions

For the naïve forecasts, the prior day's value is used for each of the following *j* forecasted days. For the HW forecasts, exponential smoothing of the prior *k* day's values is used to forecast values over the next *j* days. For the growth rate forecasts, the prior 1 day's value is used to calculate the current growth rate over the following *j* days. Then, the prior day's values for all the counties are used to calculate the growth rate for Virginia for the same *j* days. A group of *n* forecasts are generated for the county by uniformly sampling a growth rate between the county's rates and Virginia's rates. The average of the *n* forecasts is utilized as the final forecast for the county. For the MA (1), AR (1), ARMA (1, 1), and ARIMA (1, 1, 1) forecasts, the prior *k* days are given equal weighting to forecast the next *j* days.

This study only relies on the daily reported case numbers since the date of first onset within each location and does not incorporate assumptions about the basic reproductive number of COVID-19. Forecasts are influenced by the reliability of the data, the variables utilized, and the perceptions and reactions to danger and they assume the continuation of past patterns [4,31]. When exploring real-time forecasts of infectious disease models, real-time models have shown higher absolute error values, on average, than full-data models as a result of factors such as significant differences in population sizes between compared areas [32].

Statistical Analysis

We aggregated median absolute error (MdAE) and median absolute percentage error (MdAPE) variables and expressed them as medians, IQRs, and notch ranges representing 95% CIs. Notch ranges are calculated as $\pm 1.58 \times IQR/\sqrt{n}$ [33,34] as implemented using the R function geom_boxplot within the package ggplot2 (The R Foundation) [35]. Nonoverlapping comparisons of confidence intervals represent statistically significant differences [36,37] with *P* values less than .01 [38-41]. *P* values conveying significant differences between groups are calculated using Mood's median test [42,43]. MdAE

compared forecasting outcomes at shared geographic levels due to similarities in scale [44]. MdAPE compared each forecasting method's outcomes across geographic level due to differing scales [45].

We created 226,468 forecasts across the county, health district, and state levels over the period of March 7 through April 22, 2020. Due to the naïve and growth rate methods only utilizing 1-day look-backs and the ARMA and ARIMA methods requiring more than 1-day look-backs, five forecast methods exist for each comparison. Analyses were performed with R software, version 3.6.3 (The R Foundation).

Verification, Validation, and Reproducibility

The data set and code are provided in Lynch and Gore [46] and the experimental methods and steps needed for reproducibility are provided in Lynch and Gore [47]. Code inspections and unit tests were utilized for code verification [48]. MdAE and MdAPE error metrics were used for validation. A comparison of COVID-19 case count data sources found that the differences in reported case counts between *The New York Times*, *Johns Hopkins University*, and *USAFacts* do not indicate inferior or superior sources [49].

Results

Overview

Comparing all forecast methods' MdAE values across the county, health district, and state levels over the first 46 days of infection revealed that MA forecasts using 3-day look-back and 1-day forecast length achieved the lowest MdAE. This MA forecast combination was statistically significantly different in MdAE from 39 of the 59 other combinations at the county level (66 %), 53 of 59 (90%) at the health district level, and 51 of 59 (86%) at the state level. This result shows that the use of an equally weighted linear dependency with a stationary mean between the prior 3-day COVID-19 cumulative case counts, within the MA forecasts, is an effective assumption when forecasting next-day case growths for Virginia at the county, health district, and state levels. Table 1 provides the method with the lowest MdAE and the percentage of other methods from which the difference is determined to be statistically significant at the county level. Table 2 provides this information at the health district level and Table 3 provides this information at the state level.

For the methods using single-day look-back across all levels, the growth rate and naïve methods provided the lowest MdAE at the county, health district, and state levels for all forecast lengths. In general, all five methods achieved similar error confidence intervals when utilizing 1-day look-back. Only at the health district and state levels for 1-day forecast lengths was the growth rate method's difference from the other methods statistically significant, with the growth rate method performing better than all combinations at the state level.

For the methods using 7-day look-backs across all levels, the MA and AR methods were the only ones with MdAE instances that were statistically significantly lower than the other methods. The HW and ARIMA methods achieved the lowest MdAE in two instances but did not perform significantly better than the other methods in either instance. In no instance did the ARMA method obtain the lowest MdAE. The performance of the MA and AR methods supports the assumption of linear dependence between the 7-day prior days' cases and the forecast case counts. However, for the MA method the mean weighting of past values was stationary, while for the AR method it was nonstationary.

For the methods using look-backs of all prior-day case counts across all levels, the MA method achieved the lowest MdAE in all cases. This provides evidence in support of forecasting cumulative case counts using the assumption of a linear dependency and stationary mean among past values to forecast 1, 3, and 7 days when incorporating all prior cumulative cases.

Table 1. County-level median absolute error (MdAE) outcomes by forecasting method, look-back length, and forecast length.

Methods	Look-back length (days), n	Look-ahead length (days), n	df[a]	Forecasting method with lowest MdAE	Median (95% CI) (cumulative cases)	*P* value[b]	Statistically significantly lower MdAE than other methods[c], n (%)
All (N=60)	All	All	59	MA (3, 1)[d]	0.67 (0.49-0.84)	<.001	39 (66)
G1[e] (n=5)	1	1	4	Naïve	0.67 (0.43-0.90)	.09	0 (0)
G1 (n=5)	1	3	4	Naïve	1.30 (0.88-1.73)	.66	0 (0)
G1 (n=5)	1	7	4	Naïve	2.43 (1.69-3.18)	.50	0 (0)
G2[f] (n=5)	3	1	4	MA	0.67 (0.49-0.84)	.09	0 (0)
G2 (n=5)	3	3	4	MA	0.76 (0.59-0.94)	<.001	4 (100)[g]
G2 (n=5)	3	7	4	MA	1.69 (1.36-2.01)	<.001	3 (75)
G2 (n=5)	7	1	4	HW	0.91 (0.63-1.18)	.03	0 (0)
G2 (n=5)	7	3	4	MA	1.30 (0.95-1.65)	.002	1 (25)
G2 (n=5)	7	7	4	MA	2.32 (1.75-2.90)	.01	0 (0)
G2 (n=5)	All prior	1	4	MA	0.70 (0.53-0.87)	.33	0 (0)
G2 (n=5)	All prior	3	4	MA	0.83 (0.67-1.00)	<.001	4 (100)[g]
G2 (n=5)	All prior	7	4	MA	1.73 (1.36-2.10)	<.001	1 (25)

[a]Degrees of freedom represent the number of forecasting combinations minus one.

[b]*P* values were calculated for statistically significant differences in medians across groups.

[c]This was based on comparisons of notch ranges. MdAE was interpreted within geographic levels.

[d]MA: moving average; (3, 1) represents a 3-day look-back and a single-day forecast length.

[e]G1 includes naïve, MA, autoregressive (AR), growth rate, and Holt-Winters (HW) methods.

[f]G2 includes MA, AR, growth rate, HW, autoregressive moving average, and autoregressive integrated moving average methods.

[g]MA (3, 3) and MA (all prior, 3) achieved statistically significantly smaller MdAE than all four alternatives.

Table 2. Health district–level median absolute error (MdAE) outcomes by forecasting method, look-back length, and forecast length.

Methods	Look-back length (days), n	Look-ahead length (days), n	df[a]	Forecasting method with lowest MdAE	Median (95% CI) (cumulative cases)	*P* value[b]	Statistically significantly lower MdAE than other methods[c], n (%)
All (N=60)	All	All	59	MA (3, 1)[d]	3.07 (2.41-3.74)	<.001	53 (90)
G1[e] (n=5)	1	1	4	Growth rate	4.03 (3.01-5.04)	.31	1 (25)
G1 (n=5)	1	3	4	Growth rate	8.96 (6.56-11.36)	.93	0 (0)
G1 (n=5)	1	7	4	Growth rate	16.48 (11.67-21.28)	.96	0 (0)
G2[f] (n=5)	3	1	4	MA	3.07 (2.41-3.74)	.01	1 (25)
G2 (n=5)	3	3	4	MA	3.20 (2.50-3.90)	<.001	4 (100)[g]
G2 (n=5)	3	7	4	MA	7.88 (5.71-10.05)	<.001	1 (25)
G2 (n=5)	7	1	4	AR	3.57 (2.67-4.47)	.01	0 (0)
G2 (n=5)	7	3	4	MA	5.52 (3.96-7.08)	<.001	2 (50)
G2 (n=5)	7	7	4	AR	11.83 (8.16-15.49)	<.001	1 (25)
G2 (n=5)	All prior	1	4	MA	3.14 (2.47-3.80)	.04	1 (25)
G2 (n=5)	All prior	3	4	MA	3.16 (2.54-3.78)	<.001	3 (75)
G2 (n=5)	All prior	7	4	MA	7.68 (6.22-9.14)	<.001	3 (75)

[a]Degrees of freedom represent the number of forecasting combinations minus one.

[b]*P* values were calculated for statistically significant differences in medians across groups.

[c]This was based on comparisons of notch ranges. MdAE was interpreted within geographic levels.

[d]MA: moving average; (3, 1) represents a 3-day look-back and a single-day forecast length.

[e]G1 includes naïve, MA, autoregressive (AR), growth rate, and Holt-Winters (HW) methods.

[f]G2 includes MA, AR, growth rate, HW, autoregressive moving average, and autoregressive integrated moving average methods.

[g]MA (3, 3) achieved statistically significantly smaller MdAE than all four alternatives.

Table 3. State-level median absolute error (MdAE) outcomes by forecasting method, look-back length, and forecast length.

Methods	Look-back length (days), n	Look-ahead length (days), n	df[a]	Forecasting method with lowest MdAE	Median (95% CI) (cumulative cases)	*P* value[b]	Statistically significantly lower MdAE than other methods[c], n (%)
All (N=60)	All	All	59	MA (3, 1)[d]	17.43 (7.74-27.11)	<.001	51 (86)
G1[e] (n=5)	1	1	4	Growth rate	31.50 (6.11-56.89)	<.001	4 (100)[f]
G1 (n=5)	1	3	4	Growth rate	317.50 (163.15-471.85)	.94	0 (0)
G1 (n=5)	1	7	4	Growth rate	325.00 (169.49-480.51)	.18	0 (0)
G2[g] (n=5)	3	1	4	MA	17.43 (7.74-27.11)	<.001	4 (100)[f]
G2 (n=5)	3	3	4	MA	64.94 (45.93-83.96)	<.001	1 (25)
G2 (n=5)	3	7	4	MA	206.57 (148.57-264.94)	.03	1 (25)
G2 (n=5)	7	1	4	AR	69.37 (34.23-104.51)	.09	1 (25)
G2 (n=5)	7	3	4	MA	82.14 (42.83-121.47)	.02	2 (50)
G2 (n=5)	7	7	4	ARIMA	312.36 (146.54-478.17)	.012	0 (0)
G2 (n=5)	All prior	1	4	MA	25.13 (11.61-38.65)	.27	0 (0)
G2 (n=5)	All prior	3	4	MA	32.67 (21.20-44.14)	.002	2 (50)
G2 (n=5)	All prior	7	4	MA	104.85 (70.67-139.03)	.09	2 (50)

[a]Degrees of freedom represent the number of forecasting combinations minus one.

[b]*P* values were calculated for statistically significant differences in medians across groups.

[c]This was based on comparisons of notch ranges. MdAE was interpreted within geographic levels.

[d]MA: moving average; (3, 1) represents a 3-day look-back and a single-day forecast length.

[e]G1 includes naïve, MA, autoregressive (AR), growth rate, and Holt-Winters (HW) methods.

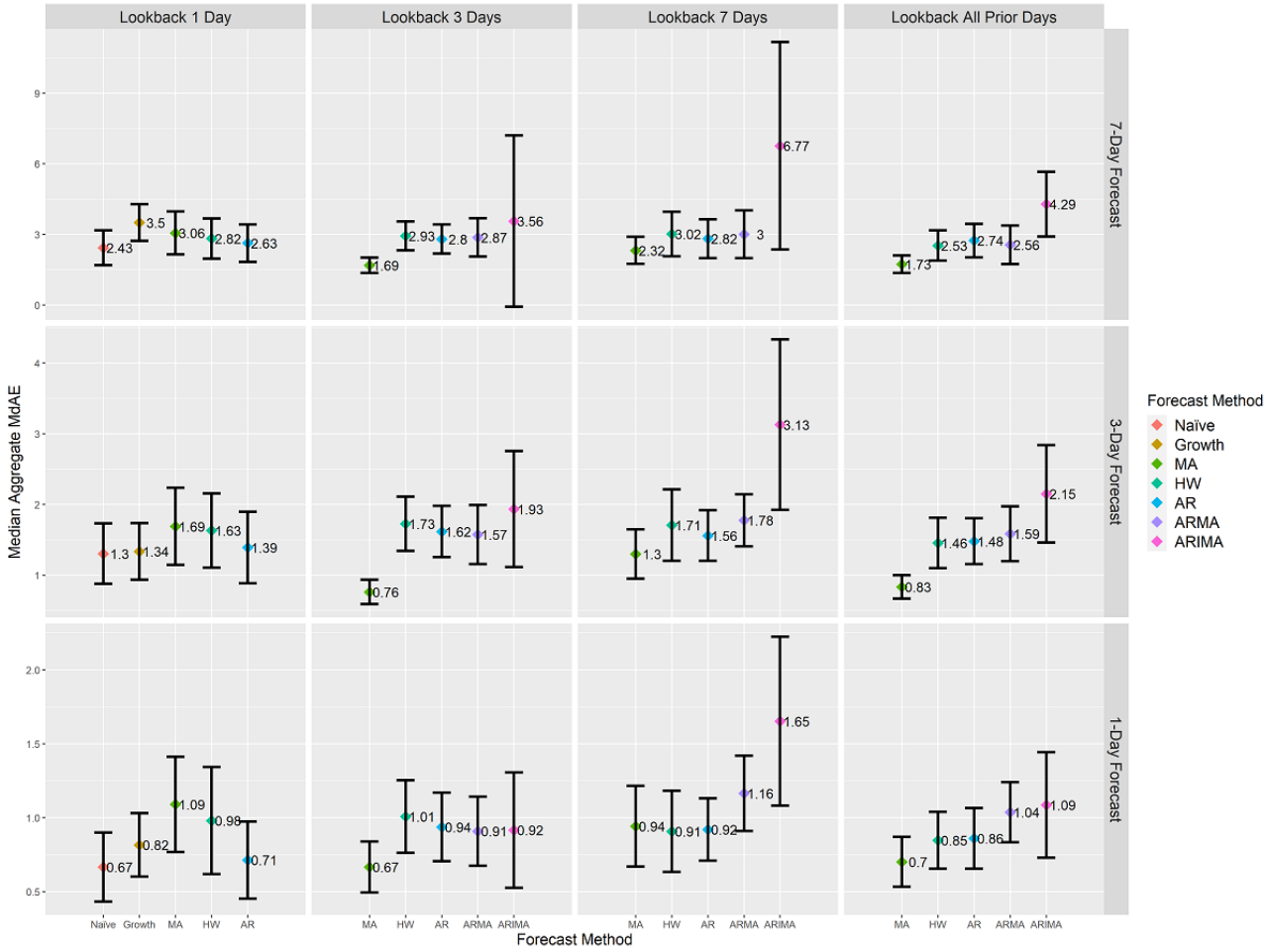
[f]Growth rate (1, 1) and MA (3, 1) achieved statistically significantly smaller MdAE than all four alternatives.

[g]G2 includes MA, AR, growth rate, HW, autoregressive moving average, and autoregressive integrated moving average (ARIMA) methods.

County-Level MdAE Results

At the county level, the MA method always achieved a lower MdAE than the ARMA method. Similarly, the ARMA method always achieved a lower MdAE than the ARIMA method. Thus, the ARIMA method's aggregated error was greater than the ARMA method's aggregated error, which was greater than the MA method's aggregated error. This indicates that the assumption of a stationary mean (ie, MA) in prior case counts is more effective than the assumption of a weakly stationary mean (ie, ARMA), which is more effective than a nonstationary mean (ie, ARIMA) when forecasting at the county level. The ARIMA method had the widest confidence interval for the median error range, indicating the least consistency in COVID-19 forecasts among these methods. Figure 2 provides the county-level MdAE outcomes for each look-back and forecast length combination. The individual results of each of the 60 forecasting combinations at the county level are provided in Multimedia Appendix 1, including median values, confidence intervals, whiskers, sample sizes, and *P* values. An interactive version of Figure 2 is provided in Multimedia Appendix 2.

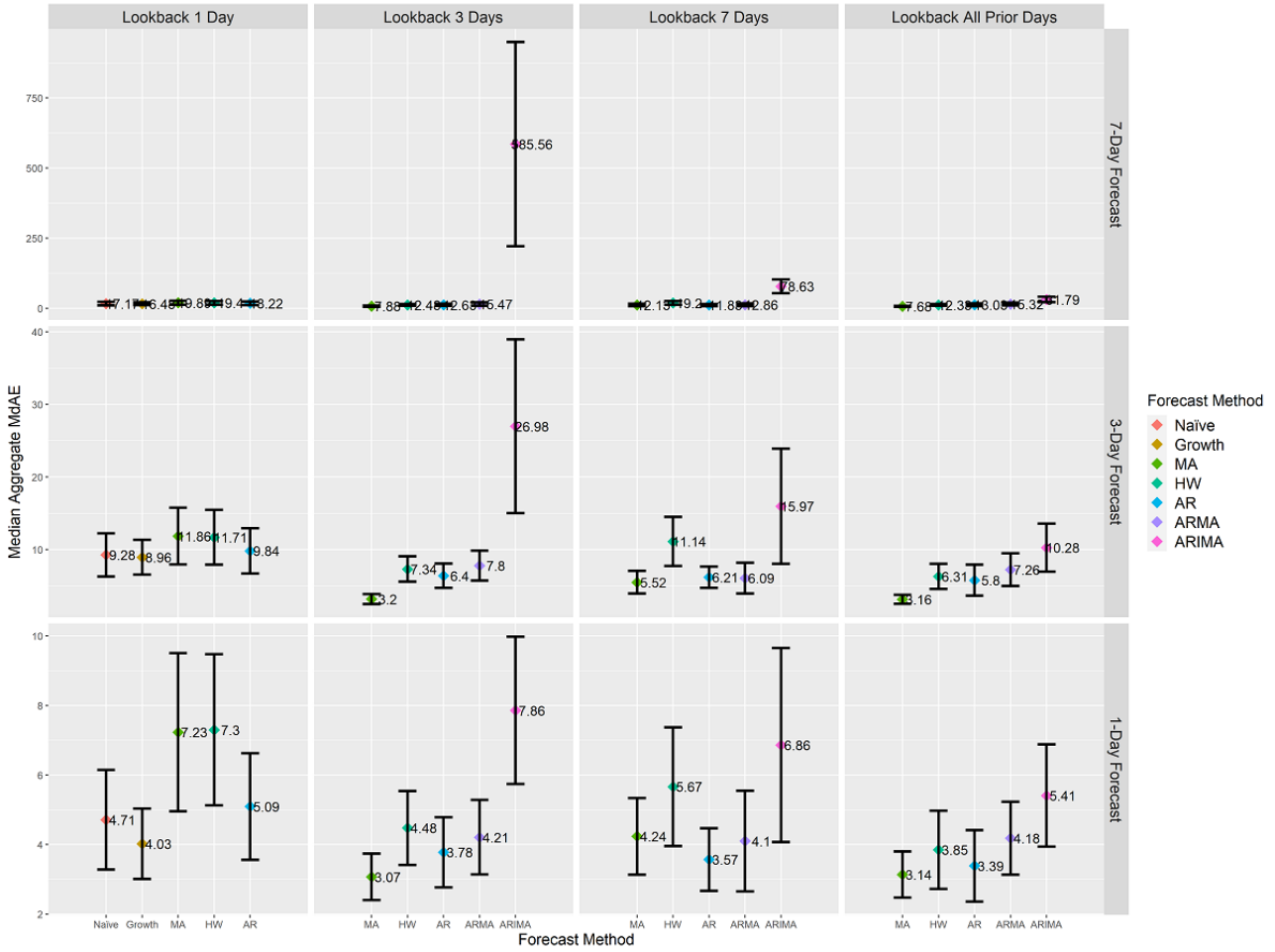
Figure 2. County-level forecasts' aggregated median MdAE values and 95% CI. CI ranges are calculated using box plot notch ranges around the median. Statistically significant differences at a *P* value of .01 are identified by nonoverlapping CI ranges of forecasting methods at each combination of forecast length and look-back length. Units are in terms of COVID-19 cumulative case counts. Y-axis scales differ on each row based on the scale of the contained data. Due to differing assumptions, five of the seven forecasting methods are present for each look-back length as indicated on the x-axis. AR: autoregressive; ARIMA: autoregressive integrated moving average; ARMA: autoregressive moving average; HW: Holt-Winters; MA: moving average; MdAE: median absolute error.



Health District–Level MdAE Results

At the health district level, the MA method always achieved lower MdAE than the ARMA method, which achieved lower MdAE than the ARIMA method. This further provided evidence that effective forecasting of cumulative COVID-19 case counts contains an assumption of stationary means in past observations. For 3-day look-back lengths with 3-day forecasts, the MA method achieved statistically significantly lower MdAE than all other methods. Figure 3 provides the MdAE at the intersection of look-back length and forecast length at the health district level. The individual results of each of the 60 forecasting combinations at the health district level are provided in Multimedia Appendix 3, including median values, confidence intervals, whiskers, sample sizes, and *P* values. An interactive version of Figure 3 is provided in Multimedia Appendix 4.

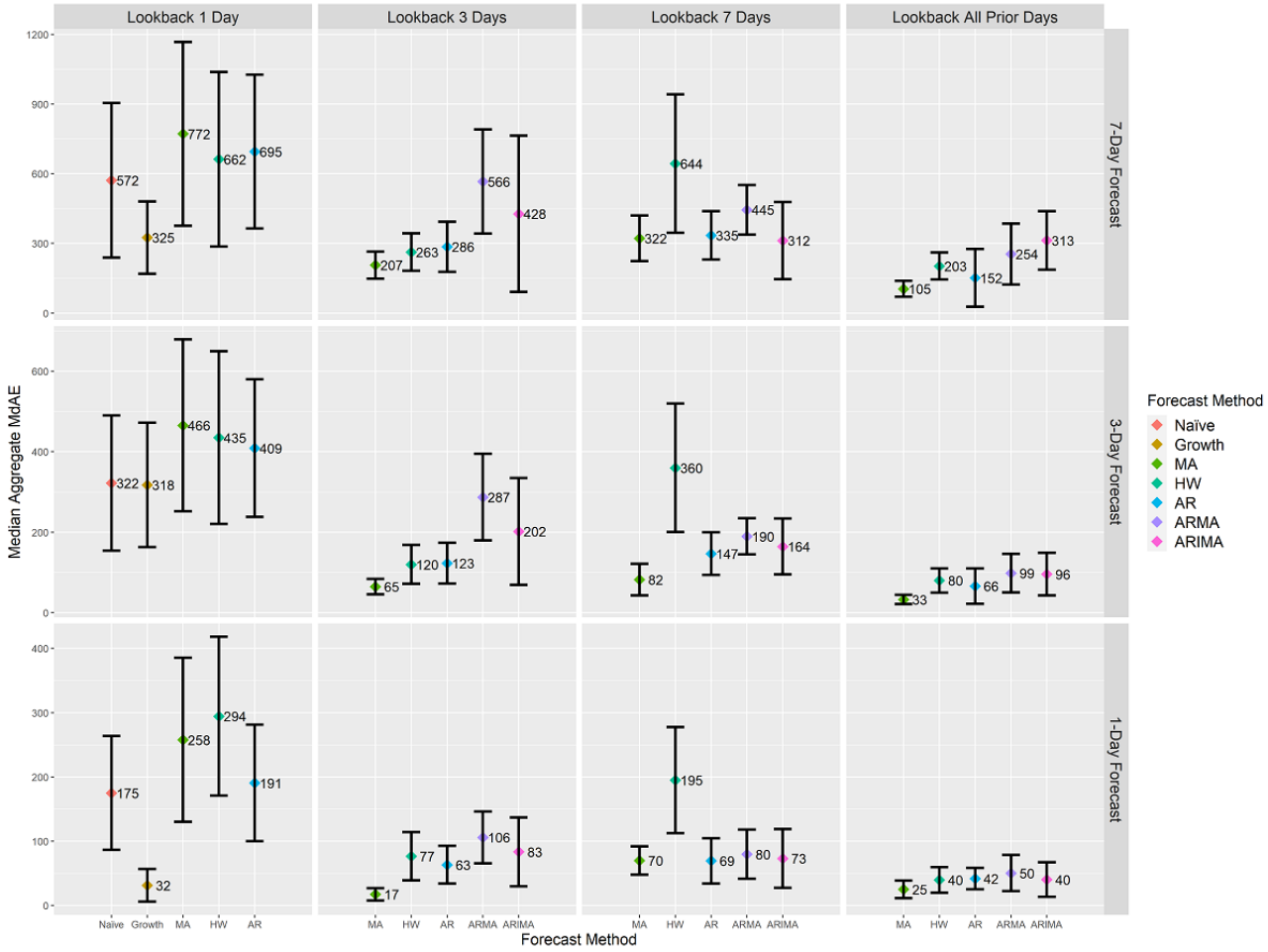
Figure 3. Health district–level forecasts' aggregated median MdAE values and 95% CI. CI ranges are calculated using box plot notch ranges around the median. Statistically significant differences at a *P* value of .01 are identified by nonoverlapping CI ranges of forecasting methods at each combination of forecast length and look-back length. Units are in terms of COVID-19 cumulative case counts. Y-axis scales differ on each row based on the scale of the contained data. Due to differing assumptions, five of the seven forecasting methods are present for each look-back length as indicated on the x-axis. AR: autoregressive; ARIMA: autoregressive integrated moving average; ARMA: autoregressive moving average; HW: Holt-Winters; MA: moving average; MdAE: median absolute error.



State-Level MdAE Results

At the state level, the growth rate method was the most effective method. In every case, it either (1) attained the lowest MdAE value compared to the other methods or (2) had the smallest notch range. The ARMA and ARIMA methods both maintained MdAE notch bands that were similar to the other methods when utilizing all prior day and 7-day look-back lengths. However, the HW method's MdAE notch bands increased as that of the ARMA and ARIMA methods decreased. These results make it unclear as to which of the assumptions related to stationary means were most effective for forecasting with the ARMA and ARIMA methods at the state level. Figure 4 provides MdAE values at the intersection of look-back length and forecast length at the state level. The individual results of each of the 60 forecasting combinations at the state level are provided in Multimedia Appendix 5, including median values, confidence intervals, whiskers, sample sizes, and *P* values. An interactive version of Figure 4 is provided in Multimedia Appendix 6.

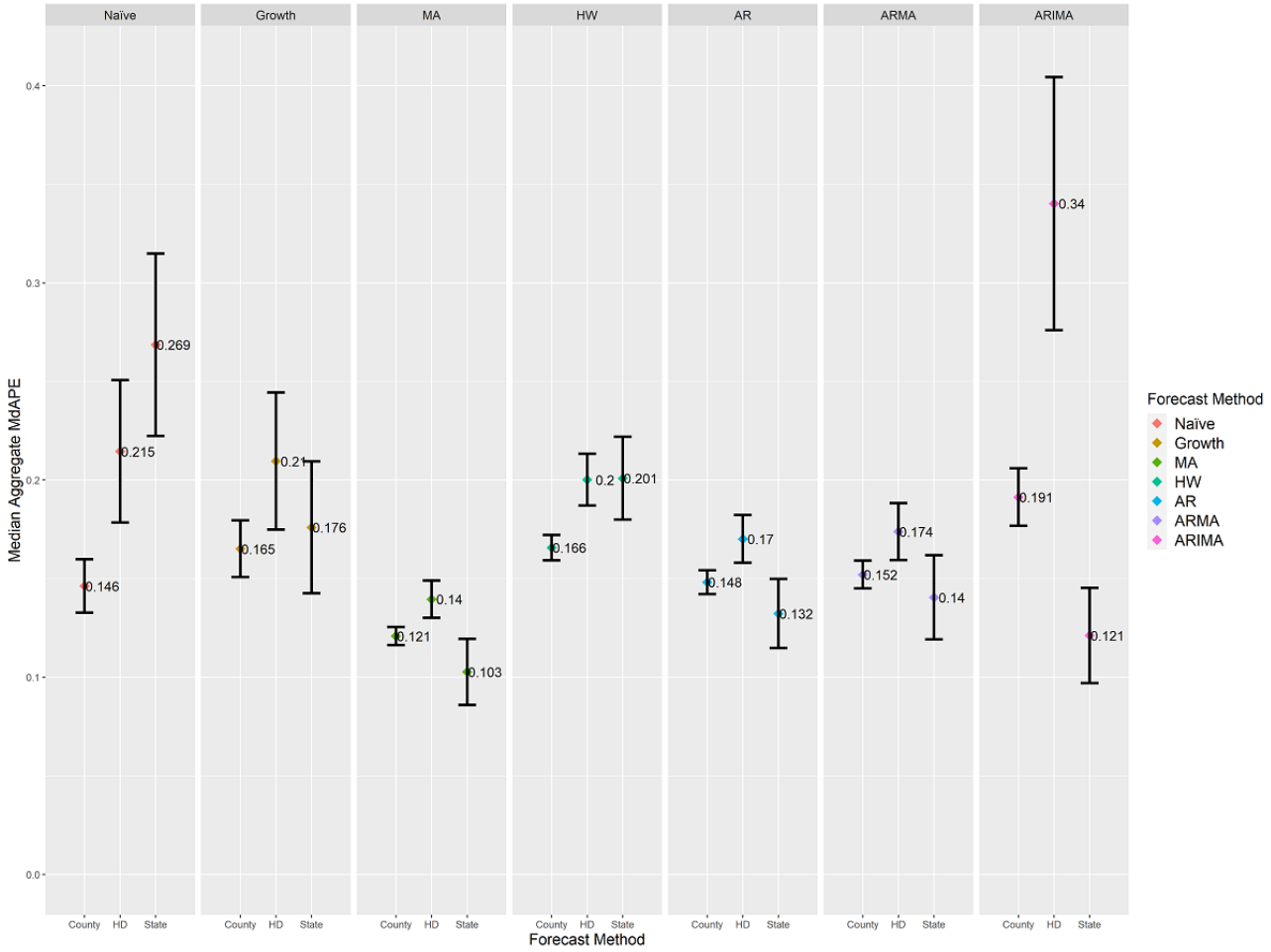
Figure 4. State-level forecasts' aggregated median MdAE values and 95% CI. CI ranges are calculated using box plot notch ranges around the median. Statistically significant differences at a *P* value of .01 are identified by nonoverlapping CI ranges of forecasting methods at each combination of forecast length and look-back length. Units are in terms of COVID-19 cumulative case counts. Y-axis scales differ on each row based on the scale of the contained data. Due to differing assumptions, five of the seven forecasting methods are present for each look-back length as indicated on the x-axis. AR: autoregressive; ARIMA: autoregressive integrated moving average; ARMA: autoregressive moving average; HW: Holt-Winters; MA: moving average; MdAE: median absolute error.



Cross-Geographic-Level MdAPE Results

MdAE reflects the scale of the data and is not appropriate for making inferences about changes in confirmed case counts between county, health district, and state levels [26,44]. Figures 2-4 convey differing scales of error values across the three levels. As a result, it was not possible to evaluate results featured in these figures against each other. To remedy this shortcoming, we applied MdAPE to identify statistically significant differences for each forecasting method individually when applied to county, health district, and state levels as provided in Figure 5. The individual results of each of the 60 forecasting combinations at the county, health district, and state levels are provided in Multimedia Appendices 7-9, including median values, confidence intervals, whiskers, sample sizes, and *P* values. An interactive version of Figure 5 is provided in Multimedia Appendix 10.

Figure 5. Aggregated median MdAPE values and 95% CI ranges at the county, health district (HD), and state levels differentiated by forecasting method. Comparing CI ranges for a forecast method across each geographic level reveals statistically significant differences in median values for the forecasting method due to geographic scale. Nonoverlapping CI ranges indicate statistically significant differences at a *P* value of .01. MdAPE provides a comparison within each forecast method separately, not a comparison across different methods. AR: autoregressive; ARIMA: autoregressive integrated moving average; ARMA: autoregressive moving average; HW: Holt-Winters; MA: moving average; MdAPE: median absolute percentage error.



Statistically significant differences were observable within a forecasting method across the county, health district, and state levels during the initial 46 days of confirmed COVID-19 case spread within Virginia. The growth rate method was the only one whose performance did not statistically significantly differ across levels; thereby, it was the only method unaffected by geographic level. The naïve method achieved a statistically significantly lower MdAPE at the county level than at the health district and state levels. The MA, HW, AR, ARMA, and ARIMA methods all contained instances of achieving statistically significantly lower MdAPE scores at the county and/or state levels than at the health district level.

Discussion

Principal Findings

Our results show the effectiveness of seven forecasting methods for the first 46 days of virus spread within Virginia at the county, health district, and state levels. In addition, a daily view of the growth rate forecast at the county level from March 7, 2020, to the present is publicly available online [50]. Tracking case and death counts yield insight into the virus's impact on a geographic region at a given point in time. Forecasts utilize the trends and averages of prior case count observations to provide expectations of case counts into the future. These forecasts keep the public informed on the state of the virus across the world and on virus levels within their own geographic areas of interest. Additionally, forecasts inform public policy for combatting the spread of the virus, supporting public health, and helping to anticipate the impacts of medical burdens across regions [51]. However, interpreting forecast outcomes requires understanding the assumptions behind the forecasting method as well as the assumptions pertaining to the geographic area and the presence of intervention strategies. Therefore, we compared the error levels pertaining to 60 forecasting combinations using the MA, AR, naïve, growth rate, HW, ARMA, and ARIMA forecasting methods. Our findings support public health with respect to forecasting by reinforcing health statistics and statistical literacy of forecasted COVID-19 outcomes.

COVID-19 cumulative case growth is such that the growth curve is exponential in the absence of preventative measures. The larger error observed in HW forecasts over MA forecasts provides support that an exponential model is not the best fit at the start of the virus spread for Virginia. The preventative measures taken by the population appear to have shifted the virus's growth behavior from exponential to linear. This finding is also supported by Lammers et al [21]. This finding supports

the idea that population interventions are effective at impacting the spread of the virus. However, as the virus continues to spread and reoccur, the inability to manufacture a vaccine for COVID-19 quickly enough to immunize the population remains a concern [52]. As such, combining short-range forecasts at the county and health district levels with targeted intervention strategies can improve planning, support, and response time. The use of rigorous government interventions may slow the rate of infections, but early detection, isolation, treatment, and adequate medical supplies are required for continued intervention against the virus [53,54].

Our cross-geographic validation checks using MdAPE indicate that the level of geographic resolution should be considered when creating forecasts of expected case counts. A forecast utilized at the state level is not likely to be as useful for determining expected growth when disaggregated across its counties during early virus onset. This results from the differing geographic assumptions present within counties or health districts when compared to the state. This finding is consistent with the literature reporting that case growths vary across countries and across states [14]. Variations result from factors such as population behaviors in response to the pandemic, implemented policy interventions, and population densities [7,9,10,15]. Furthermore, since the growth rate method did not produce statistically significant MdAPE differences across geographic levels, it may be a good choice for decision makers whose region does not match the county, health district, or state levels.

To identify a best option among our tested combinations, we compared MdAE ranges against each other within each geographic tier. The MA method using a 3-day look-back length and a single-day forecast length provided the smallest error (ie, lowest MdAE) at the county level (median 0.67, 95% CI 0.49-0.84; $P<.001$), the health district level (median 3.07, 95% CI 2.41-3.74; $P<.001$), and the state level (median 17.43, 95% CI 7.74-27.11; $P<.001$). Compared to the other forecasting combinations, the MA method's confidence intervals statistically significantly differed from 39 out of 59 alternatives (66%; county level) to 51 out of 59 alternatives (86%; state level) to 53 out of 59 alternatives (90%; health district level) at a P value level of .01. When relying on only the prior day's case counts, the growth rate method stood out as the best option at the health district and state levels; however, the naïve, growth rate, HW, MA, and AR methods performed similarly well at the county level.

When utilizing 3 or more days of prior observations, a diverse range of options is available. For next-day forecasts, there was no method that performed better at a level that was statistically significant among the five options. For 3-day forecasts, the MA method was statistically significantly better than 25% to 100% of the other four options in all cases. For 7-day forecasts, the MA method performed statistically significantly better than 25% to 75% of the other four options when using a 3-day look-back or an all prior–day look-back. When using a 7-day look-back, the AR method performed the best and its difference from the ARIMA method was statistically significant. The ARMA and ARIMA methods achieved the lowest error in any of the combinations.

These findings support the assumption of stationarity within the mean of the prior days' cumulative case counts. This is reflected in how well the MA method performed and how poorly the ARIMA method performed at forecasting cumulative case counts. Rarely do the ARIMA or ARMA methods achieve lower error values than any other combination. This reflects the idea that the assumption of stationary means of past observations is a more effective representation of cumulative COVID-19 growth than assumptions of weak stationarity or nonstationarity. The need to apply a differencing step to remove nonstationarity using the ARIMA method is not present within the data during this period. Additionally, placing extra weight on the recent past does not improve forecasting during this period, as the HW and AR methods were consistently less effective than the MA method. These findings suggest that the ARIMA and ARMA methods are unlikely to be good fits and should not be used to forecast case counts during early onset within areas that have only a few weeks of historical data collected, whose residents are aware of the existence of the virus and are engaging in preventative behaviors, and that contain similar population densities to Virginia.

Several studies utilized forecasts to estimate case fatality and recovery ratios, epidemiological parameters, and transmission dynamics based on data from the start of the outbreak [55,56]. Studies also support the idea that epidemiological differences contribute to variations in the severity of the contracted disease [2,57]. Based on historical similarities to previous influenza strains, social distancing can potentially reduce transmission of the virus; however, the effectiveness may vary alongside changes in seasonal factors in travel as well as between tropical and temperate climates [58,59]. Distancing may be especially beneficial in rural areas, where fewer hospitals and health care facilities exist, by emphasizing strategies oriented toward specific population age groups [60].

The results of this study can be expanded to include areas' demographic characteristics, geographic characteristics, and preventative measures to strive for more accurate forecast models. A recent study found COVID-19 growth to strongly correlate with population density, percent of the population living in rural areas, and yearly flu vaccination rate [14]. Exploring forecast behaviors of areas sharing these traits may further reduce forecasting error and reveal subgroupings of viable forecasting method options. Additionally, forecasting models can be paired with mortality models [61] to gain better estimates of infection forecasts per demographic characteristics. Predictive methods derived from search engines' data [62] can also be incorporated within forecasting methods. This would provide a way to connect forecasts with human search behaviors based on the frequency of searched terms identified in relation to COVID-19 prevention and recovery. Forecasting models can be paired with models of local medical burden and pandemic preparedness [51] for more detailed representations of expected medical strains and greater flexibility in testing preventative measures.

Stay-at-home orders have been successful as intervention strategies to slow the spread of the virus. However, the impact caused by a neighboring area's removal of mobility restrictions needs greater exploration [10]. The secondary effects of the

starting or ending of proximity-based prevention methods on neighboring areas can help assess the potential impact of a mitigation strategy. For identified hot spots, the Centers for Disease Control and Prevention provides outreach to local officials and helps in identifying adapted interventions for the local area [7]. Rapid identification and timeliness of response are critical, especially if the impact of interventions can take up to 3 weeks to be effective [15]. Reliable forecasting can aid in the identification of emerging hot spots and support timely response. To this end, increased knowledge of forecasting characteristics based on geographic level, demographic characteristics, population density characteristics, and population behaviors can help reveal the primary drivers of upcoming cases. This knowledge can be leveraged to inform early, targeted interventions or to provide risk updates to targeted populations within an area. People could then modify their mobility and social decisions themselves in a timely manner separate from population mandated measures.

Limitations

Here we discuss internal and external validity threats as well as other limitations that affected our work. Internal validity threats arise when factors affect the dependent variables without the researchers' knowledge. It is possible that some implementation flaws could have affected our modeling results or the ensuing data analysis. However, the algorithms in our source code were (1) built on established libraries, (2) passed several internal code reviews, and (3) are publicly accessible, along with the data and results. Threats to external validity occur when the results of our analysis and our simulation cannot be generalized. Our results are limited to Virginia, from March to April 2020 with respect to COVID-19 cases reported by *The New York Times*. Our results are not immediately generalizable to (1) different infectious diseases, (2) other COVID-19 data sets, (3) different periods of time, or (4) different geographic areas.

Several other assumptions and limitations pertain to this study. Seven forecasting methods with differing baseline assumptions were evaluated with respect to how well they forecast the early growth of COVID-19 cases within Virginia; however, numerous additional forecasting methods exist with different combinations of assumptions that can also be explored with respect to this pandemic. Conclusions should not be drawn about the effectiveness of these findings for forecast lengths greater than 7 days, as the appropriateness of underlying assumptions, such as stationarity of prior days' values, would need to be re-evaluated. Larger median error values of the 7-day forecasts, versus their 1- and 3-day counterparts, were observable, further supporting the need for evaluation of forecast assumptions pertaining to the characteristics of COVID-19 beyond 7-day forecast lengths. The selected forecasting methods assumed that policies and population behaviors remained unchanged during the forecast periods; therefore, the forecasts do not account for future starts or ends of policies, such as stay-at-home orders or return-to-work dates.

Finally, our results do not reflect how the spread of the virus would occur for locations experiencing first contact with the virus without yet having public awareness of the existence of the virus. These findings are applicable under the assumption that the general population was already aware of the presence of COVID-19. At the starting point of Virginia cases, the local population was already aware that cases had reached the United States, the virus had been classified as a pandemic, and the virus was receiving major media attention. Additionally, the governor had issued an executive order declaring a state of emergency due to COVID-19 on March 12, 2020 [10,63]. This provided 47 days for Virginians to prepare and modify their standard movement and interaction behaviors as they deemed necessary for their own safety. As a result, these results were not captured in the same context as the period of time when areas within the United States were first impacted by the virus (ie, areas of Washington, California, and New York).

Authors' Contributions

All authors contributed to the conceptualization of the study and the data collection. CL conducted the literature review, constructed the experimental design, and drafted the manuscript. RG developed the procedure for the growth rate forecasts and led code development. All authors contributed to data analysis, data interpretation, figure creation, and revisions and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
County-level median absolute error (MdAE) outcomes for each forecasting combination.
[XLSX File (Microsoft Excel File), 16 KB - jmir_v23i3e24925_app1.xlsx]

Multimedia Appendix 2
Interactive plot conveying county-level median absolute error (MdAE) outcomes for each forecasting combination.
[ZIP File (Zip Archive), 1038 KB - jmir_v23i3e24925_app2.zip]

Multimedia Appendix 3
Health district–level median absolute error (MdAE) outcomes for each forecasting combination.

[XLSX File (Microsoft Excel File), 16 KB - jmir_v23i3e24925_app3.xlsx]

Multimedia Appendix 4
Interactive plot conveying health district–level median absolute error (MdAE) outcomes for each forecasting combination.
[ZIP File (Zip Archive), 1038 KB - jmir_v23i3e24925_app4.zip]

Multimedia Appendix 5
State-level median absolute error (MdAE) outcomes for each forecasting combination.
[XLSX File (Microsoft Excel File), 16 KB - jmir_v23i3e24925_app5.xlsx]

Multimedia Appendix 6
Interactive plot conveying state-level median absolute error (MdAE) outcomes for each forecasting combination.
[ZIP File (Zip Archive), 1038 KB - jmir_v23i3e24925_app6.zip]

Multimedia Appendix 7
County-level median absolute percentage error (MdAPE) outcomes for each forecasting combination.
[XLSX File (Microsoft Excel File), 16 KB - jmir_v23i3e24925_app7.xlsx]

Multimedia Appendix 8
Health district–level median absolute percentage error (MdAPE) outcomes for each forecasting combination.
[XLSX File (Microsoft Excel File), 17 KB - jmir_v23i3e24925_app8.xlsx]

Multimedia Appendix 9
State-level median absolute percentage error (MdAPE) outcomes for each forecasting combination.
[XLSX File (Microsoft Excel File), 17 KB - jmir_v23i3e24925_app9.xlsx]

Multimedia Appendix 10
Interactive figure conveying county-, health district–, and state-level median absolute percentage error (MdAPE) outcomes for each forecasting combination.
[ZIP File (Zip Archive), 1034 KB - jmir_v23i3e24925_app10.zip]

References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
2. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
3. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman J, et al. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infect Dis Model 2020;5:256-263 [FREE Full text] [doi: 10.1016/j.idm.2020.02.002] [Medline: 32110742]
4. Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. PLoS One 2020;15(3):e0231236 [FREE Full text] [doi: 10.1371/journal.pone.0231236] [Medline: 32231392]
5. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
6. Singh RK, Rani M, Bhagavathula AS, Sah R, Rodriguez-Morales AJ, Kalita H, et al. Prediction of the COVID-19 pandemic for the top 15 affected countries: Advanced autoregressive integrated moving average (ARIMA) model. JMIR Public Health Surveill 2020 May 13;6(2):e19115 [FREE Full text] [doi: 10.2196/19115] [Medline: 32391801]
7. Oster AM, Kang GJ, Cha AE, Beresovsky V, Rose CE, Rainisch G, et al. Trends in number and distribution of COVID-19 hotspot counties - United States, March 8-July 15, 2020. MMWR Morb Mortal Wkly Rep 2020 Aug 21;69(33):1127-1132 [FREE Full text] [doi: 10.15585/mmwr.mm6933e2] [Medline: 32817606]
8. Yu P, Zhu J, Zhang Z, Han Y. A familial cluster of infection associated with the 2019 novel coronavirus indicating possible person-to-person transmission during the incubation period. J Infect Dis 2020 May 11;221(11):1757-1761 [FREE Full text] [doi: 10.1093/infdis/jiaa077] [Medline: 32067043]
9. Gallaway MS, Rigler J, Robinson S, Herrick K, Livar E, Komatsu KK, et al. Trends in COVID-19 incidence after implementation of mitigation measures - Arizona, January 22-August 7, 2020. MMWR Morb Mortal Wkly Rep 2020 Oct 09;69(40):1460-1463 [FREE Full text] [doi: 10.15585/mmwr.mm6940e3] [Medline: 33031366]

10. Moreland A, Herlihy C, Tynan MA, Sunshine G, McCord RF, Hilton C, CDC Public Health Law Program, CDC COVID-19 Response Team, Mitigation Policy Analysis Unit. Timing of state and territorial COVID-19 stay-at-home orders and changes in population movement - United States, March 1-May 31, 2020. MMWR Morb Mortal Wkly Rep 2020 Sep 04;69(35):1198-1203 [FREE Full text] [doi: 10.15585/mmwr.mm6935a2] [Medline: 32881851]
11. Siedner MJ, Harling G, Reynolds Z, Gilbert RF, Haneuse S, Venkataramani AS, et al. Social distancing to slow the US COVID-19 epidemic: Longitudinal pretest–posttest comparison group study. PLoS Med 2020 Aug 11;17(8):e1003244. [doi: 10.1371/journal.pmed.1003244]
12. Lyu W, Wehby GL. Shelter-in-place orders reduced COVID-19 mortality and reduced the rate of growth in hospitalizations. Health Aff (Millwood) 2020 Sep;39(9):1615-1623. [doi: 10.1377/hlthaff.2020.00719] [Medline: 32644825]
13. Turk PJ, Chou S, Kowalkowski MA, Palmer PP, Priem JS, Spencer MD, et al. Modeling COVID-19 latent prevalence to assess a public health intervention at a state and regional scale: Retrospective cohort study. JMIR Public Health Surveill 2020 Jun 19;6(2):e19353 [FREE Full text] [doi: 10.2196/19353] [Medline: 32427104]
14. White ER, Hébert-Dufresne L. State-level variation of initial COVID-19 dynamics in the United States. PLoS One 2020;15(10):e0240648 [FREE Full text] [doi: 10.1371/journal.pone.0240648] [Medline: 33048967]
15. Li Y, Campbell H, Kulkarni D, Harpur A, Nundy M, Wang X, et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: A modelling study across 131 countries. Lancet Infect Dis 2021 Feb;21(2):193-202. [doi: 10.1016/s1473-3099(20)30785-4]
16. Wang Y, Zhang Y, Bennell K, White DK, Wei J, Wu Z, et al. Physical distancing measures and walking activity in middle-aged and older residents in Changsha, China, during the COVID-19 epidemic period: Longitudinal observational study. J Med Internet Res 2020 Oct 26;22(10):e21632 [FREE Full text] [doi: 10.2196/21632] [Medline: 33027035]
17. Liu D, Clemente L, Poirier C, Ding X, Chinazzi M, Davis J, et al. Real-time forecasting of the COVID-19 outbreak in Chinese provinces: Machine learning approach using novel digital data and estimates from mechanistic models. J Med Internet Res 2020 Aug 17;22(8):e20285 [FREE Full text] [doi: 10.2196/20285] [Medline: 32730217]
18. Armstrong JS. The forecasting dictionary. In: Armstrong JS, editor. Principles of Forecasting: A Handbook for Researchers and Practitioners. Boston, MA: Springer; 2001:761-819.
19. Tovissodé CF, Lokonon BE, Glèlè Kakaï R. On the use of growth models to understand epidemic outbreaks with application to COVID-19 data. PLoS One 2020;15(10):e0240578 [FREE Full text] [doi: 10.1371/journal.pone.0240578] [Medline: 33079964]
20. Singh RK, Drews M, De La Sen M, Kumar M, Singh SS, Pandey AK, et al. Short-term statistical forecasts of COVID-19 infections in India. IEEE Access 2020;8:186932-186938. [doi: 10.1109/access.2020.3029614]
21. Lammers J, Crusius J, Gast A. Correcting misperceptions of exponential coronavirus growth increases support for social distancing. Proc Natl Acad Sci U S A 2020 Jul 14;117(28):16264-16266 [FREE Full text] [doi: 10.1073/pnas.2006048117] [Medline: 32581118]
22. Pourghasemi HR, Pouyan S, Farajzadeh Z, Sadhasivam N, Heidari B, Babaei S, et al. Assessment of the outbreak risk, mapping and infection behavior of COVID-19: Application of the autoregressive integrated-moving average (ARIMA) and polynomial models. PLoS One 2020;15(7):e0236238 [FREE Full text] [doi: 10.1371/journal.pone.0236238] [Medline: 32722716]
23. Singh RK, Rani M, Bhagavathula AS, Sah R, Rodriguez-Morales AJ, Kalita H, et al. Prediction of the COVID-19 pandemic for the top 15 affected countries: Advanced autoregressive integrated moving average (ARIMA) model. JMIR Public Health Surveill 2020 May 13;6(2):e19115 [FREE Full text] [doi: 10.2196/19115] [Medline: 32391801]
24. Hyndman RJ, Athanasopoulos G. Forecasting: Principles and Practice. 2nd edition. Melbourne, Australia: OTexts; 2018.
25. Armstrong JS. Evaluating forecasting methods. In: Armstrong JS, editor. Principles of Forecasting: A Handbook for Researchers and Practitioners. Boston, MA: Springer; 2001:443-472.
26. Armstrong J, Collopy F. Error measures for generalizing about forecasting methods: Empirical comparisons. Int J Forecast 1992 Jun;8(1):69-80. [doi: 10.1016/0169-2070(92)90008-w]
27. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. Psychol Sci Public Interest 2007 Nov;8(2):53-96. [doi: 10.1111/j.1539-6053.2008.00033.x] [Medline: 26161749]
28. Wallman KK. Enhancing statistical literacy: Enriching our society. J Am Stat Assoc 1993 Mar;88(421):1. [doi: 10.2307/2290686]
29. Amidon TR, Nielsen AC, Pflugfelder EH, Richards DP, Stephens SH. Visual risk literacy in “flatten the curve” COVID-19 visualizations. J Bus Tech Commun 2020 Oct 01;35(1):101-109. [doi: 10.1177/1050651920963439]
30. Heesterbeek H, Anderson RM, Andreasen V, Bansal S, De Angelis D, Dye C, Isaac Newton Institute IDD Collaboration. Modeling infectious disease dynamics in the complex landscape of global health. Science 2015 Mar 13;347(6227):aaa4339 [FREE Full text] [doi: 10.1126/science.aaa4339] [Medline: 25766240]
31. Tizzoni M, Bajardi P, Poletto C, Ramasco JJ, Balcan D, Gonçalves B, et al. Real-time numerical forecast of global epidemic spreading: Case study of 2009 A/H1N1pdm. BMC Med 2012 Dec 13;10:165 [FREE Full text] [doi: 10.1186/1741-7015-10-165] [Medline: 23237460]

32. Reich NG, Lauer SA, Sakrejda K, Iamsirithaworn S, Hinjoy S, Suangtho P, et al. Challenges in real-time prediction of infectious disease: A case study of dengue in Thailand. PLoS Negl Trop Dis 2016 Jun;10(6):e0004761 [FREE Full text] [doi: 10.1371/journal.pntd.0004761] [Medline: 27304062]
33. Nuzzo RL. The box plots alternative for visualizing quantitative data. PM R 2016 Mar;8(3):268-272. [doi: 10.1016/j.pmrj.2016.02.001] [Medline: 26892802]
34. Chambers JM, Cleveland WS, Kleiner B, Tukey PA. Graphical Methods for Data Analysis. London, UK: Chapman and Hall/Cole Publishing Co; 1983.
35. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2nd edition. Cham, Switzerland: Springer; 2016.
36. Krzywinski M, Altman N. Visualizing samples with box plots. Nat Methods 2014 Feb;11(2):119-120. [doi: 10.1038/nmeth.2813] [Medline: 24645192]
37. McGill R, Tukey JW, Larsen WA. Variations of box plots. Am Stat 1978 Feb;32(1):12. [doi: 10.2307/2683468]
38. Altman DG, Bland JM. How to obtain the P value from a confidence interval. BMJ 2011;343:d2304. [doi: 10.1136/bmj.d2304] [Medline: 22803193]
39. Cumming G, Finch S. Inference by eye: Confidence intervals and how to read pictures of data. Am Psychol 2005;60(2):170-180. [doi: 10.1037/0003-066x.60.2.170]
40. Cumming G. Inference by eye: Reading the overlap of independent confidence intervals. Stat Med 2009 Jan 30;28(2):205-220. [doi: 10.1002/sim.3471] [Medline: 18991332]
41. Cumming G, Finch S. A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. Educ Psychol Meas 2001;61(4):532-574. [doi: 10.1177/0013164401614002]
42. Divine GW, Norton HJ, Barón AE, Juarez-Colunga E. The Wilcoxon–Mann–Whitney procedure fails as a test of medians. Am Stat 2018 Mar 15;72(3):278-286. [doi: 10.1080/00031305.2017.1305291]
43. Brown GW, Mood AM. On median tests for linear hypotheses. In: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, CA: University of California Press; 1951 Presented at: Second Berkeley Symposium on Mathematical Statistics and Probability; July 31-August 12, 1950; Berkeley, CA p. 159-166 URL: https://projecteuclid.org/ebooks/bsmsp/Proceedings%20of%20the%20Second%20Berkeley%20Symposium%20on%20Mathematical%20Statistics%20and%20Probability/chapter/On%20Median%20Tests%20for%20Linear%20Hypotheses/bsmsp/1200500226
44. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. Int J Forecast 2006 Oct;22(4):679-688. [doi: 10.1016/j.ijforecast.2006.03.001]
45. Shcherbakov MV, Brebels A, Shcherbakova NL, Tyukov AP, Janovsky TA, Kamaev V. A survey of forecast error measures. World Appl Sci J 2013;24:171-176 [FREE Full text] [doi: 10.5829/idosi.wasj.2013.24.itmies.80032]
46. Lynch CJ, Gore R. Short-range early phase COVID-19 forecasting R-project and data, V2. Mendeley Data. 2020 Dec 15. URL: https://data.mendeley.com/datasets/cytrb8p42g/2 [accessed 2021-03-02]
47. Lynch CJ, Gore R. Application of one-, three-, and seven-day forecasts during early onset on the COVID-19 epidemic dataset using moving average, autoregressive, autoregressive moving average, autoregressive integrated moving average, and naïve forecasting methods. Data Brief 2021 Apr;35:106759 [FREE Full text] [doi: 10.1016/j.dib.2021.106759] [Medline: 33521186]
48. Lynch CJ, Diallo SY, Kavak H, Padilla JJ. A content analysis-based approach to explore simulation verification and identify its current challenges. PLoS One 2020;15(5):e0232929 [FREE Full text] [doi: 10.1371/journal.pone.0232929] [Medline: 32401795]
49. Wang G, Gu Z, Li X, Yu S, Kim M, Wang Y, et al. Comparing and integrating US COVID-19 data from multiple sources with anomaly detection and repairing. arXiv. Preprint posted online on November 28, 2020 [FREE Full text]
50. Gore R, Lynch CJ. Virginia County COVID-19 Daily Case Total Forecaster. Suffolk, VA: Virginia Modeling, Analysis and Simulation Center (VMASC), Old Dominion University; 2020. URL: https://vmasc.shinyapps.io/va-county-covid-forecast/ [accessed 2021-03-02]
51. Sha D, Miao X, Lan H, Stewart K, Ruan S, Tian Y, et al. Spatiotemporal analysis of medical resource deficiencies in the US under COVID-19 pandemic. PLoS One 2020;15(10):e0240348 [FREE Full text] [doi: 10.1371/journal.pone.0240348] [Medline: 33052956]
52. Lipsitch M, Finelli L, Heffernan RT, Leung GM, Redd SC, 2009 H1N1 Surveillance Group. Improving the evidence base for decision making during a pandemic: The example of 2009 influenza A/H1N1. Biosecur Bioterror 2011 Jun;9(2):89-115 [FREE Full text] [doi: 10.1089/bsp.2011.0007] [Medline: 21612363]
53. Fang Y, Nie Y, Penny M. Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis. J Med Virol 2020 Jun;92(6):645-659 [FREE Full text] [doi: 10.1002/jmv.25750] [Medline: 32141624]
54. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
55. Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLoS One 2020;15(3):e0230405 [FREE Full text] [doi: 10.1371/journal.pone.0230405] [Medline: 32231374]

56. Zhang J, Litvinova M, Wang W, Wang Y, Deng X, Chen X, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: A descriptive and modelling study. Lancet Infect Dis 2020 Jul;20(7):793-802 [FREE Full text] [doi: 10.1016/S1473-3099(20)30230-9] [Medline: 32247326]
57. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: A model-based analysis. Lancet Infect Dis 2020 Jun;20(6):669-677 [FREE Full text] [doi: 10.1016/S1473-3099(20)30243-7] [Medline: 32240634]
58. Cobey S. Modeling infectious disease dynamics. Science 2020 May 15;368(6492):713-714. [doi: 10.1126/science.abb5659] [Medline: 32332062]
59. Shaman J, Kohn M. Absolute humidity modulates influenza survival, transmission, and seasonality. Proc Natl Acad Sci U S A 2009 Mar 03;106(9):3243-3248 [FREE Full text] [doi: 10.1073/pnas.0806852106] [Medline: 19204283]
60. Ranscombe P. Rural areas at risk during COVID-19 pandemic. Lancet Infect Dis 2020 May;20(5):545 [FREE Full text] [doi: 10.1016/S1473-3099(20)30301-7] [Medline: 32311327]
61. Yang W, Kandula S, Huynh M, Greene S, Van Wye G, Li W, et al. Estimating the infection-fatality risk of SARS-CoV-2 in New York City during the spring 2020 pandemic wave: A model-based analysis. Lancet Infect Dis 2021 Feb;21(2):203-212 [FREE Full text] [doi: 10.1016/S1473-3099(20)30769-6] [Medline: 33091374]
62. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, Kalhori SRN. Predicting COVID-19 incidence through analysis of Google Trends data in Iran: Data mining and deep learning pilot study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18828 [FREE Full text] [doi: 10.2196/18828] [Medline: 32234709]
63. Northam RS, Thomasson K. Executive Order Number Fifty-One: Declaration of a State of Emergency Due to Novel Coronavirus (COVID-19). Richmond, VA: Commonwealth of Virginia, Office of the Governor; 2020 Mar 12. URL: https://www.fmcsa.dot.gov/sites/fmcsa.dot.gov/files/2020-03/VA%20Governor%20Northam%203%2012%202020%20Declaration-of-a-State-of-Emergency-Due-to-Novel-Coronavirus-%28COVID-19%29.pdf [accessed 2021-03-02]

Abbreviations

AR: autoregressive
ARIMA: autoregressive integrated moving average
ARMA: autoregressive moving average
HW: Holt-Winters
MA: moving average
MdAE: median absolute error
MdAPE: median absolute percentage error

Edited by G Eysenbach, C Basch; submitted 10.10.20; peer-reviewed by D Valdez, Z Ren; comments to author 27.10.20; revised version received 07.12.20; accepted 22.02.21; published 23.03.21.

Please cite as:
Lynch CJ, Gore R
Short-Range Forecasting of COVID-19 During Early Onset at County, Health District, and State Geographic Levels Using Seven Methods: Comparative Forecasting Study
J Med Internet Res 2021;23(3):e24925
URL: https://www.jmir.org/2021/3/e24925
doi:10.2196/24925
PMID:33621186

Original Paper

Echo Chamber Effect in Rumor Rebuttal Discussions About COVID-19 in China: Social Media Content and Network Analysis Study

Dandan Wang[1,2,3], PhD; Yuxing Qian[1,2,3], MSc

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

[2]Center for Studies of Information Resources, Wuhan University, Wuhan, China

[3]Big Data Institute, Wuhan University, Wuhan, China

Corresponding Author:
Dandan Wang, PhD
School of Information Management
Wuhan University
No.483, Bayi Road, Hongshan District
Wuhan, 430072
China
Phone: 86 13871241431
Email: dandanw@whu.edu.cn

Abstract

Background: The dissemination of rumor rebuttal content on social media is vital for rumor control and disease containment during public health crises. Previous research on the effectiveness of rumor rebuttal, to a certain extent, ignored or simplified the structure of dissemination networks and users' cognition as well as decision-making and interaction behaviors.

Objective: This study aimed to roughly evaluate the effectiveness of rumor rebuttal; dig deeply into the attitude-based echo chamber effect on users' responses to rumor rebuttal under multiple topics on Weibo, a Chinese social media platform, in the early stage of the COVID-19 epidemic; and evaluate the echo chamber's impact on the information characteristics of user interaction content.

Methods: We used Sina Weibo's application programming interface to crawl rumor rebuttal content related to COVID-19 from 10 AM on January 23, 2020, to midnight on April 8, 2020. Using content analysis, sentiment analysis, social network analysis, and statistical analysis, we first analyzed whether and to what extent there was an echo chamber effect on the shaping of individuals' attitudes when retweeting or commenting on others' tweets. Then, we tested the heterogeneity of attitude distribution within communities and the homophily of interactions between communities. Based on the results at user and community levels, we made comprehensive judgments. Finally, we examined users' interaction content from three dimensions—sentiment expression, information seeking and sharing, and civility—to test the impact of the echo chamber effect.

Results: Our results indicated that the retweeting mechanism played an essential role in promoting polarization, and the commenting mechanism played a role in consensus building. Our results showed that there might not be a significant echo chamber effect on community interactions and verified that, compared to like-minded interactions, cross-cutting interactions contained significantly more negative sentiment, information seeking and sharing, and incivility. We found that online users' information-seeking behavior was accompanied by incivility, and information-sharing behavior was accompanied by more negative sentiment, which was often accompanied by incivility.

Conclusions: Our findings revealed the existence and degree of an echo chamber effect from multiple dimensions, such as topic, interaction mechanism, and interaction level, and its impact on interaction content. Based on these findings, we provide several suggestions for preventing or alleviating group polarization to achieve better rumor rebuttal.

(J Med Internet Res 2021;23(3):e27009) doi:10.2196/27009

KEYWORDS

rumor rebuttal; infodemiology; infodemic; infoveillance; echo chamber effect; attitude; COVID-19; Weibo

Introduction

Background

In the early stage of the COVID-19 crisis, social science played an essential role in the containment of the disease [1]. The core part of the public health defense strategy lay in geographical social distancing [2]. However, the unique characteristics of social media, such as diversification of information, liberalization of expression, and efficient transmission speed [3], had facilitated the propagation of a large number of rumors related to the spreading and blocking of the COVID-19 epidemic (eg, "The Air Force of the Central Theater District spreads disinfectant powder over Wuhan," "Smoking and drinking can kill the new coronavirus," and "From March 16th, citizen travel will be normalized"), resulting in public pressure and panic [4,5]. The probability of the public adopting constructive behavior (eg, maintaining personal hand hygiene and avoiding group aggregation) or disruptive behavior (eg, panic buying and adopting unproven treatments) largely depended on whether managers conveyed necessary deterministic information in a timely manner through online means to clarify rumors [6]. Evaluating the public's reaction or attitude toward rumor rebuttal could help confirm the effectiveness of rumor rebuttal so as to guide the public to implement health decisions and actions based on the correct information [6,7].

While social media serves as a breeding ground for rumors, it is also directly used for rumor management [8-10]. However, the lack of fact verification of rumor rebuttal released by the public, as well as the national media or government organizations fabricating incorrect stories to conceal facts, lead to more intense controversy about rumor rebuttal, greatly dispelling the effectiveness of rumor rebuttal during public events [11]. In addition, the emergence of social media as an information dissemination channel shortens the distance from content producers to consumers and profoundly changes the way users obtain information, debate, and shape their attitudes [12]. Firstly, the information filtering mechanism based on algorithm recommendation technology mediates and promotes content promotion by considering users' preferences and attitudes. Secondly, affected by individual and social factors, such as selective psychological mechanisms, group pressures, and social network circles, online users tend to choose the information that conforms to their belief system and ignore information that does not conform to their beliefs, eventually forming echo chambers, reflected as homophily-based communities of like-minded people that strengthen their shared narrative [13,14]. High segregation and clustering within this homophily-based community may increase the polarization of attitudes toward issues or events of public concern [15]. In the case of a user embedded in a community in which most users disagree with the rumor rebuttal, he or she is very likely to obey group norms, ignore the few voices within community that advocate refuting rumors, and even cut off communication with other communities holding rumor rebuttal views. Then, the scale of anti–rumor rebuttal groups continues to grow, and opposition groups are highly isolated. As a result, no matter how many repeated, mild attacks of rumor rebuttal are released, users' ordinary beliefs (ie, agree or disagree with rumor rebuttal) are difficult to change [16,17]. Therefore, it is critical to explore how rumor rebuttal diffuses in a contemporary media environment where users can easily filter and choose their information sources.

Goals of This Study

First, this research aimed to analyze the distribution of users' attitudes (ie, agree, disagree, query, or unknown) toward rumor rebuttal related to COVID-19 on Chinese social media (ie, Weibo, the Chinese equivalent of Twitter [15]) to check its effectiveness. Second, we focused on determining whether and to what extent (ie, polarization based on users' attitudes or even based on communities' mixed attitudes) the echo chamber effect existed in the process of shaping users' attitudes when retweeting or commenting on rumor rebuttal. Last but not least, we needed to evaluate the impact that the echo chamber effect had on the characteristics of interactive content. Specifically, using a combination of manual content analysis, automatic text analysis, social network analysis, and statistical analysis, we paid attention to retweeting and commenting networks at both the user and community levels, quantified the homophily based on the attitude distribution of nodes in networks, and checked the user interaction content from three dimensions: sentiment expression, information seeking and sharing, and civility. Generally, by understanding the community structure of the dissemination network of rumor rebuttal and the interactive nature of users within and among communities, we could comprehensively reveal the function and impact of echo chambers on guiding the public's cognition as well as their decision-making and interactive behaviors.

Prior Work

Rumor Rebuttal Dissemination and Echo Chamber Effect

Rumor spread refers to information being widely spread during uncertain or dangerous situations to alleviate fear and anxiety [18,19], while rumor rebuttal refers to the use of information to effectively combat rumors [20-23]. Research studies on the factors that influence the effectiveness of rumor rebuttal focus on the characteristics of the source subjects [10,24-27], content [28,29], and dissemination channels [30,31]. To some extent, these research studies ignore the individual differences in information receivers' responses to rumor rebuttal and the context of the communities in which communicators are embedded. Limited research studies have pointed out that individuals with different knowledge reserves, interests, and values have different perceived credibility toward rumor rebuttal [32,33], and social identity plays a vital role in the transmission cascade of rumor rebuttal [34]. Individuals develop social identity by establishing cognitive and emotional connections with social groups, organizations, or other social entities, and they have a sense of belonging to entities with common beliefs [35]. This social identity is accompanied by confirmation bias; namely, individuals tend to shape their own attitudes to be in line with their prior attitudes as well as according to the standards that they share with the people around them in order to enhance their identity [34]. Such communities of judgment provide the basis for whether individuals view the decisions of their in-group as legitimate [34,36], eventually facilitating the

emergence of echo chambers; namely, a kind of situation or circumstance where they tend to share ideas, information, or opinions with the same values, while barely considering alternative opinions [13,16]. From the perspective of social networks, the existence of clusters is understood as the evidence of echo chambers, in which nodes tend to preferentially connect to nodes in clusters and form homogeneous communities [37]. Homophily refers to the principle that contact between similar people occurs at a higher probability than between dissimilar people [38].

Based on users' different interaction behaviors in different situations, different conclusions about the existence or exposure degree of the echo chamber effect are obtained. By analyzing political rumor rebuttal on Twitter during the 2012 US presidential election campaign, Shin et al [39] found that, within retweeting networks, rumor refuters neither formed a sizable community nor exhibited a partisan structure; that is, rejecters who supported rumor rebuttal about Barack Obama also engaged in debunking rumors about Mitt Romney—the two candidates were in competition with each other. In addition, Zollo et al [16] discovered the echo chamber effect in users' interactions toward rumor rebuttal of unverified conspiracy information on Facebook: two well-formed and highly segregated communities existed around conspiracy and anticonspiracy topics (ie, users mainly liked or commented on only one category). Other research studies related to the echo chamber effect mainly focused on common and controversial public social issues, such as food safety [15], public advocacy [15,40], and climate change [41]. Less attention has been paid to the specific transmission process of rumor rebuttal, and even less research has been done to explore the public's selective acceptance behavior of rumor rebuttal during public health emergencies. Relevant research conclusions need to be supplemented and verified.

Retweeting and commenting serve as evidence of the effectiveness of information dissemination strategies [10,42] or evidence of the significance of the content that was retweeted or commented on [43]. What's more, retweets and comments can reflect public attitudes [15,40,41]. In addition, the nature of the online community may vary depending on the topic being discussed [38,41]. Based on the above trends and findings, we proposed the following research questions. Research Question 1 asks, "What was the distribution of attitudes (ie, agree, disagree, query, or unknown) toward rumor rebuttal under different topics related to COVID-19, and how effective was the rumor rebuttal?" Research Question 2 asks, "Based on the interactive mechanism of retweeting and commenting on Weibo, did the echo chamber effect exist in attitudes toward rumor rebuttal under different topics related to COVID-19? If so, to what extent?"

Sentiment, Information, Civility, and the Echo Chamber Effect

Informal networks on social media are used for different purposes, including sentiment expression, information seeking and sharing, and collective sensemaking [43]. Previous research studies on public sentiment during major public health emergencies found that people usually exhibited negative sentiments, such as panic, anxiety, anger, sadness, and disgust [44,45]. After rumors were refuted, public sentiment usually changed from negative to positive [46,47]. However, Zollo et al [16] found that negative sentiments dominated in the comments on Facebook that refuted conspiracy theories, regardless of users' polarization. In addition, social identity theory proposed that identification with the entity involves positive feelings of sympathy about the entity and, conversely, that disidentification entails negative feelings of dislike or even hate [48]. In other words, the longer the discussions between polarized communities with opposite attitudes, the more negativity was found overall [49,50].

In addition to sharing personal experiences, cracking jokes, or expressing concern, people also use social media to ask questions or find and share knowledge [51]. On the one hand, affected populations actively seek explanations to reduce uncertainty in times of crisis [52]. On the other hand, the research on Twitter messages has found that over half of the posts were information related and contained links to websites [51,53].

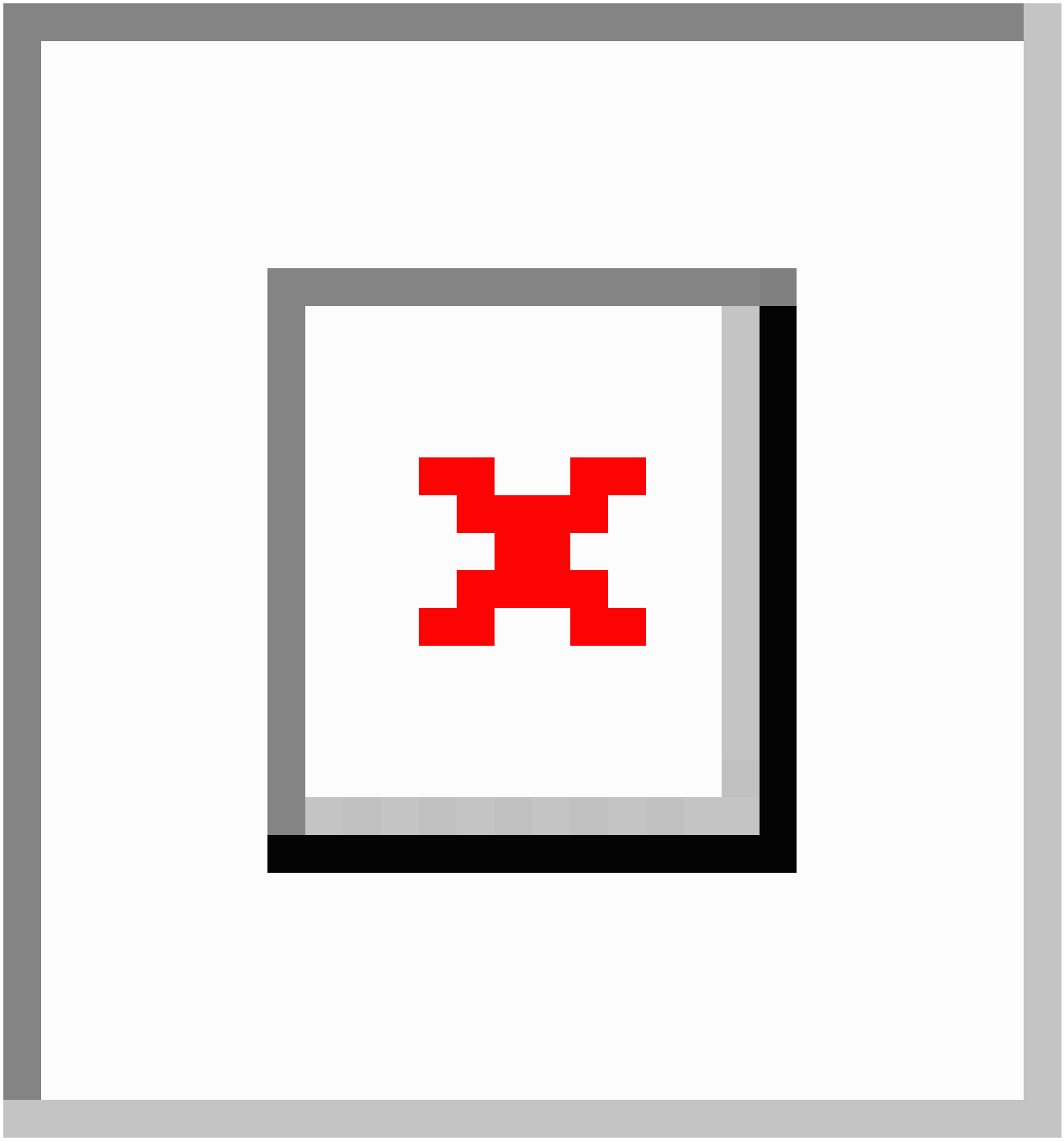
Rowe [54] pointed out that the anonymity of the digital environment reduced the quality of online communication. When users participated in highly controversial debates, uncivil, discriminatory, hateful, and other remarks were more likely to surface [55]. Chen et al [56] argued that members with different attitudes might frequently retweet each other's tweets, but meaningful or rational conversations seldom occur. In order to study the relationship between different types of user interactions and the sentiment-, information-, and civility-based tweets, we established the following research question. Research Question 3 asks, "Were the tweets from different kinds of interactions among users different in the distribution of sentiment tendency, information seeking or sharing, and civility?"

Berger [57] claimed that users experiencing amusement were more willing to share information. Wolleb et al [49] found that angry people who tended to exercise less critical judgment and relied more on stereotypes were more likely to seek out information confirming their prior attitudes to contribute to the echo chamber. Moreover, anger was also associated with incivility and hostility, which might increase distrust and polarization [58]. To study the relationship among sentiment, information, and civility, we put forward the following research question. Research Question 4 asks, "Were there any correlations when users expressed sentiment or sought or shared information when making uncivil remarks in rumor rebuttal discussions on Weibo?"

Methods

The research design for this study is outlined in Figure 1.

Figure 1. Research design. ANOVA: analysis of variance.



Data Collection and Preprocessing

As human-to-human transmission has been confirmed [59], to prevent further spread of COVID-19 from its source, the city's entire transportation system was prohibited from entering and leaving Wuhan starting from 10 AM on January 23, 2020, followed by the whole of Hubei province a day later. It was not until midnight on April 8, 2020, that Wuhan was unsealed to transportation. This period was the early stage of the COVID-19 epidemic. The high risk and uncertainty of the emerging infectious disease were likely to cause widespread public concern. Its suddenness and the insufficiency of the official response inevitably caused an information vacuum [60], which provided a breeding ground for rumors [46,61]. Due to geographical social distancing, rumor resolution mainly relied on online rebuttal. Therefore, this study first used Sina Weibo's application programming interface [10] to crawl the original tweets containing the keywords "novel coronavirus (新冠)/COVID-19" and "rumor rebuttal (辟谣)" from 10 AM on January 23, 2020, to midnight on April 8, 2020; the posters' information was also included. Next, we examined the list of retweets and comments of each original tweet to obtain the retweets and comments as well as the corresponding user information. We initially obtained 3446 original tweets, 23,858 retweets, and 12,740 comments on tweets.

Next, we performed data preprocessing. We deleted original tweets whose number of retweets or comments were below 20 to exclude low-impact samples [15]; we also deleted tweets that

contained the above keywords but had nothing to do with the content in order to exclude noise from the data (eg, "#COVID-19#Rumor rebuttal is important") [1]. Uncertain or false original rumor rebuttal tweets were also deleted. In addition to the excluded original tweets, we removed their retweets and comments. The final experimental data included 55 original rumor rebuttal tweets corresponding to 51 users, 18,118 retweets corresponding to 16,707 users, and 6064 comments corresponding to 4863 users.

Content Analysis and Sentiment Computing

Firstly, we invited two trained professionals to create the topic categories for the 55 original rumor rebuttal tweets, referring to Chen's topic classification rules for rumor rebuttal about the COVID-19 epidemic [30]. After adding and deleting certain topics, reviewing the topics, and eliminating disagreements during the actual coding process for the corpus, the topic categories were determined as shown in Table 1.

Table 1. Topic categories of original rumor rebuttal tweets.

Topic category	Explanation
Virus	The pathological characteristics of the virus, the name of the virus, etc
Contagion	The mode of disease transmission, the route of disease transmission, etc
Prevention	Disease prevention measures, related knowledge, etc
Patients	The physical health of the patients, the mental health of the patients, etc
Sequelae	The performance of sequelae in the recovered population
Epidemic situation	The spread of the epidemic in various regions
Domestic government countermeasures	Response measures of Chinese government departments at all levels
Other domestic countermeasures	Response measures of enterprises and organizations in China apart from the government
Foreign countermeasures	Response measures of countries outside China
Other	Other unimportant topics that were not equivalent to the above topics

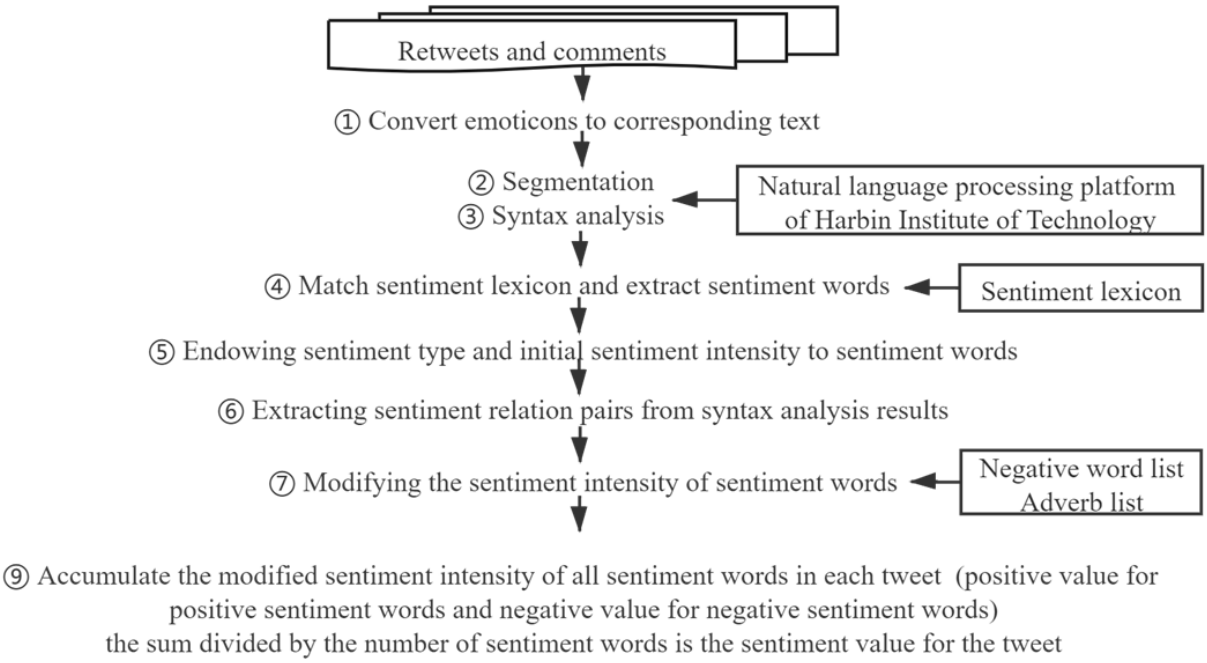
Secondly, based on the number of retweets and comments and the number of users participating in retweeting and commenting under each topic, we selected the top number of topics (N) with the highest attention [10]. To answer Research Questions 1 and 2, we divided the attitudes of all the original posters, the retweeting users, and the commenting users who participated in the discussion under these topics into four categories: (1) agree—users agreed with the rumor rebuttal, (2) disagree—users disagreed with the rumor rebuttal, (3) query—users queried the rumor rebuttal, and (4) unknown—users had no clear attitude [62]. Attitudes of all the original posters were marked as *agree*. For retweeting users, we first determined the user's attitude as expressed in a single retweet; we then comprehensively considered all of his or her retweets and selected the attitude that was most frequently expressed as the user's attitude (ie, *agree*, *disagree*, or *query*). In the case of attitudes with equal frequencies, the attitude expressed by the latest retweet prevailed. The user's attitude was coded as *unknown* if none of their retweets expressed his or her attitude clearly. The two coders coded 10% of the sample data and conducted intercoder reliability tests [63]. After eliminating differences and reaching agreement through discussion, they marked the remaining samples. By the same rule, they categorized the retweeting and commenting users, and intercoder reliability was calculated (retweeting users: κ=0.889; commenting users: κ=0.961). The high number of users who retweeted or commented on and agreed with the rumor rebuttal post served as a sign of effective refutation [10,64].

To analyze the content-related characteristics of retweets and comments, we coded them based on the dimensions of information (ie, seeking, sharing, and no information) and civility (ie, civility and incivility). The tweets that were abusive, threatening, or prejudiced against others or that were harmful to national laws and unity were classified as uncivil. The two dimensions were independent of each other. Intercoder reliability tests based on 10% of the sample data obtained κ values of 0.965 and 0.967 for information and civility of retweets, respectively, and 0.958 and 0.947 for information and civility of comments, respectively. All of the values were acceptable, suggesting that the results of the classification were robust.

Lastly, we adopted the sentiment analysis method based on the dictionary used by An and Ou [65], scored the sentiment intensity of sentiment-related words—scores of 1, 3, 5, 7, and 9, where 1 represented the lowest intensity and 9 represented the highest intensity—and considered the modification of negative words and adverbs on sentiment-related words. See equations 1 and 2 for details.

Sensibility(Pair) was the modified sentiment intensity of the sentiment-related word. *Sensibility(W)* was the initial sentiment intensity of the sentiment-related word. $Value_{adv}$ was the intensity of the adverb (continuous value between 0 and 2) that modified the sentiment-related word. *n* was the number of negative words that modified the sentiment-related word. The sentiment scores of each retweet and comment were calculated (see equation 3; *m* was the number of sentiment-related words in each tweet). If the score was greater than 0, it was regarded as positive, and if the score was less than 0, it was regarded as negative; otherwise, it was neutral. Specific steps are shown in Figure 2.

Figure 2. Sentiment computing.



Social Network Construction and Visualization

User-Level Interaction Networks

To explore the structure of the user interaction network and its homophily at individual level, we established two user-level interaction networks based on retweeting and commenting behaviors under each topic (2 × N in total). In retweeting networks, if user *i* retweeted a tweet from user *j*, then there was an edge from *i* to *j*. In commenting networks, if user *i* commented on a tweet from user *j*, then there was an edge from *i* to *j*. Both retweeting and commenting networks were directed and weighted, and the weight of the edge was determined by the number of interactions between users. Networks were constructed using Python's NetworkX package (Python Software Foundation) [66] to obtain the detailed features of topology structure and the homophily based on the users' attitudes. Furthermore, we used Gephi's Fruchterman Reinhold layout algorithm to visualize the connectivity and homophily of user networks [67].

Community-Level Interaction Networks

To explore the structure of users' community networks and its homophily at the community level, in each user-level interaction network, we used Gephi's community detection algorithm to divide each user into the corresponding community, where users interacted more frequently with each other than they did with others outside, based on the network topology and independent of the user's attitude [68]. In 2 × N community-level interaction networks, each node represented a community and edges represented the remaining interactions between users who belonged to different communities after deleting the interactions within communities. Where *query* and *unknown* attitudes were equal to 0, *agree* was equal to 1, and *disagree* was equal to –1, we summarized and averaged the attitude scores of all users in each community, which represented the community's attitude score [15,41].

A was the observed frequency of members holding the a*gree* attitude and *D* was the observed frequency of members holding the *disagree* attitude within this community. The 2 × N directed weighted networks were constructed using Python's NetworkX package [66] to obtain the detailed characteristics of topology structure and the homophily based on the community's attitude score. Furthermore, we used Gephi's Force Atlas layout algorithm to visualize the connectivity and homophily of community networks [67].

Echo Chamber Effect Measurement

Echo Chamber Effect Based on User Level

Referring to Wang and Song [15], Tsai et al [40], and Williams et al [41], in each user-level interaction network, we counted the connection of nodes with the same or different attitudes and regarded the high connection frequency between similar nodes and/or low connection frequency between dissimilar nodes as evidence of homophily.

Echo Chamber Effect Based on Community Level

On the one hand, we analyzed heterogeneity within communities [15], which was defined as the balance between users holding *agree* and *disagree* attitudes within each community node and was measured as follows:

A was the observed frequency of members holding *agree* attitudes and *D* was the observed frequency of members holding *disagree* attitudes within this community. *H* gave values on a

linear scale, ranging from perfect homogeneity (H=0, members holding only *agree* or only *disagree* attitudes) to perfect heterogeneity (H=1, equal proportions of members holding *agree* or *disagree* attitudes).

In addition, we also used Python's NetworkX package [66] to calculate the assortativity coefficient r (–1 to 1) of each network based on the attitude scores of the community nodes and the interactions among communities. The coefficient measured the homophily of the network based on the node-level attribute (ie, the community's attitude score), which essentially referred to the Pearson correlation of behaviors between the linked nodes [69-71]. If r>0, the node generally tended to connect to other nodes with similar properties and the network was called an assortative network; a larger r meant more prominent assortativity. If r≤0, this did not hold [15].

Statistical Analysis

The Mann-Whitney U nonparametric test is mainly used to test whether there is a significant difference between the averages of two groups of samples [40,72]. One-way analysis of variance (ANOVA) is used to infer significant differences among three or more independent groups' averages of a variable [15,73]. The chi-square test can be used for comparison of multiple rates or constituent ratios [74]. To answer Research Question 3, we used one-way ANOVA to compare the sentiment scores of retweets of like-minded, cross-cutting, unclear user interactions; we used the chi-square test to compare the proportion of retweets that contained information seeking, information sharing, and no information as well as civil and uncivil retweets of like-minded, cross-cutting, unclear user interactions. For commenting, we made the same comparison. To answer Research Question 4, we used one-way ANOVA to compare the sentiment scores of retweets that contained information seeking, information sharing, and no information; we used the Mann-Whitney U nonparametric test to compare the sentiment scores of retweets that were civil and uncivil; and we used the chi-square test to compare the proportion of civil and uncivil retweets that contained information seeking, information sharing, and no information. For commenting, we made the same comparison.

Results

Descriptive Statistics

The topic coding results for 55 original tweets are shown in Table 2. Apart from *others*, the top six topics that were widely retweeted and commented on by users were *epidemic situation*, *foreign countermeasures*, *prevention*, *virus*, *patients*, and *domestic government countermeasures*. The following research took the data corresponding to these six topics as the research object, including 49 original tweets published by 45 users, 10,845 retweets published by 9511 users, and 5305 comments published by 4264 users.

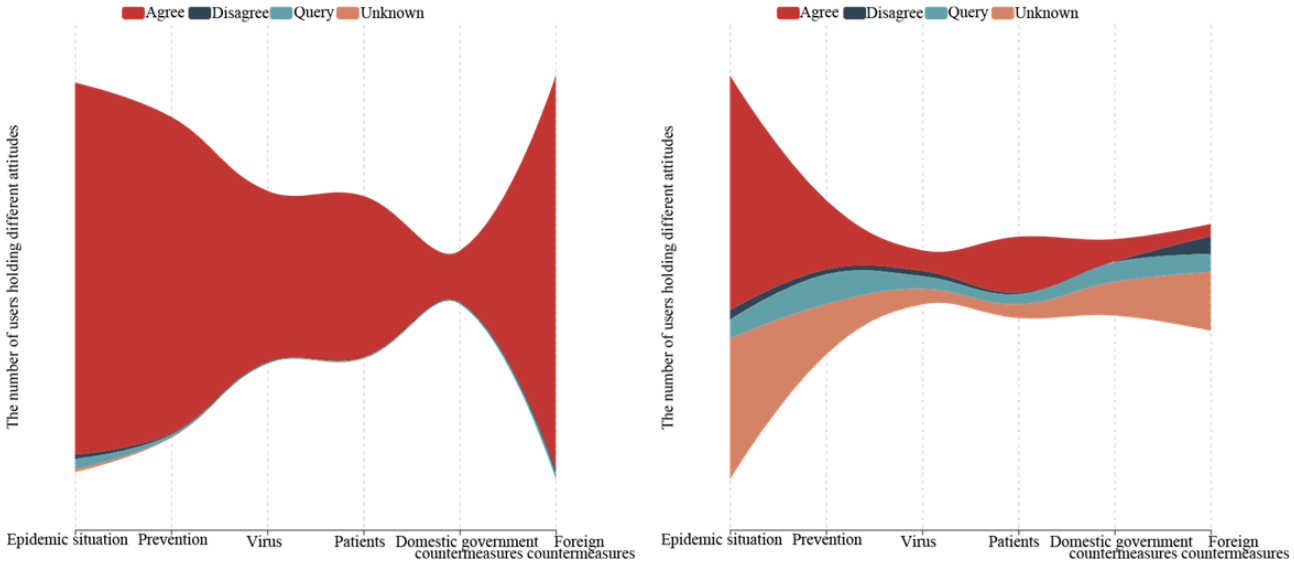
In Figure 3, the different colors of the river branches represent different attitudes, and the widths of the river branches represent the number of users holding the corresponding attitudes. Most of the users who retweeted rumor rebuttal tweets displayed the *agree* attitude. Few users who retweeted rumor rebuttal tweets about the *epidemic situation* showed the *query* attitude. However, the users who commented on rumor rebuttal tweets had more diversified attitudes under various topics, with the majority of individuals holding *query* or *unknown* attitudes. Under the topics of *epidemic situation* and *foreign countermeasures*, especially, more users showed *disagree* or *query* attitudes.

Table 2. Topic distribution of original rumor rebuttal tweets and their retweets and comments.

Topic	Original tweets, n (%)[a]		Retweets, n (%)[a]		Comments, n (%)[a]	
	Users (n=51)	Tweets (n=55)	Users (n=16,707)	Retweets (n=18,118)	Users (n=4863)	Comments (n=6064)
Virus	3 (6)	3 (5)	1204 (7.2)	1207 (6.7)	266 (5.5)	355 (5.8)
Contagion	3 (6)	3 (5)	230 (1.4)	233 (1.3)	233 (4.8)	343 (5.7)
Prevention	6 (12)	8 (15)	2235 (13.4)	2471 (13.6)	759 (15.6)	896 (14.8)
Patients	4 (8)	4 (7)	1132 (6.8)	1137 (6.3)	399 (8.2)	414 (6.8)
Treatment	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Sequelae	1 (2)	1 (2)	242 (1.4)	242 (1.3)	29 (0.6)	30 (0.5)
Epidemic situation	25 (49)	25 (46)	2693 (16.1)	2835 (15.6)	1980 (40.7)	2603 (42.9)
Domestic government countermeasures	5 (10)	5 (9)	364 (2.2)	372 (2.1)	374 (7.7)	462 (7.6)
Other domestic countermeasures	1 (2)	1 (2)	25 (0.1)	25 (0.1)	30 (0.6)	35 (0.6)
Foreign countermeasures	4 (8)	4 (7)	2813 (16.8)	2823 (15.6)	526 (10.8)	575 (9.5)
Others	1 (2)	1 (2)	6773 (40.5)	6773 (37.4)	309 (6.4)	351 (5.8)

[a]Percentages may add up to greater than 100% because one user may post multiple original tweets among which each tweet belongs to a specific topic.

Figure 3. Distribution of attitudes of users who retweeted (left) or commented on (right) the original rumor rebuttal tweets under the top six topics.

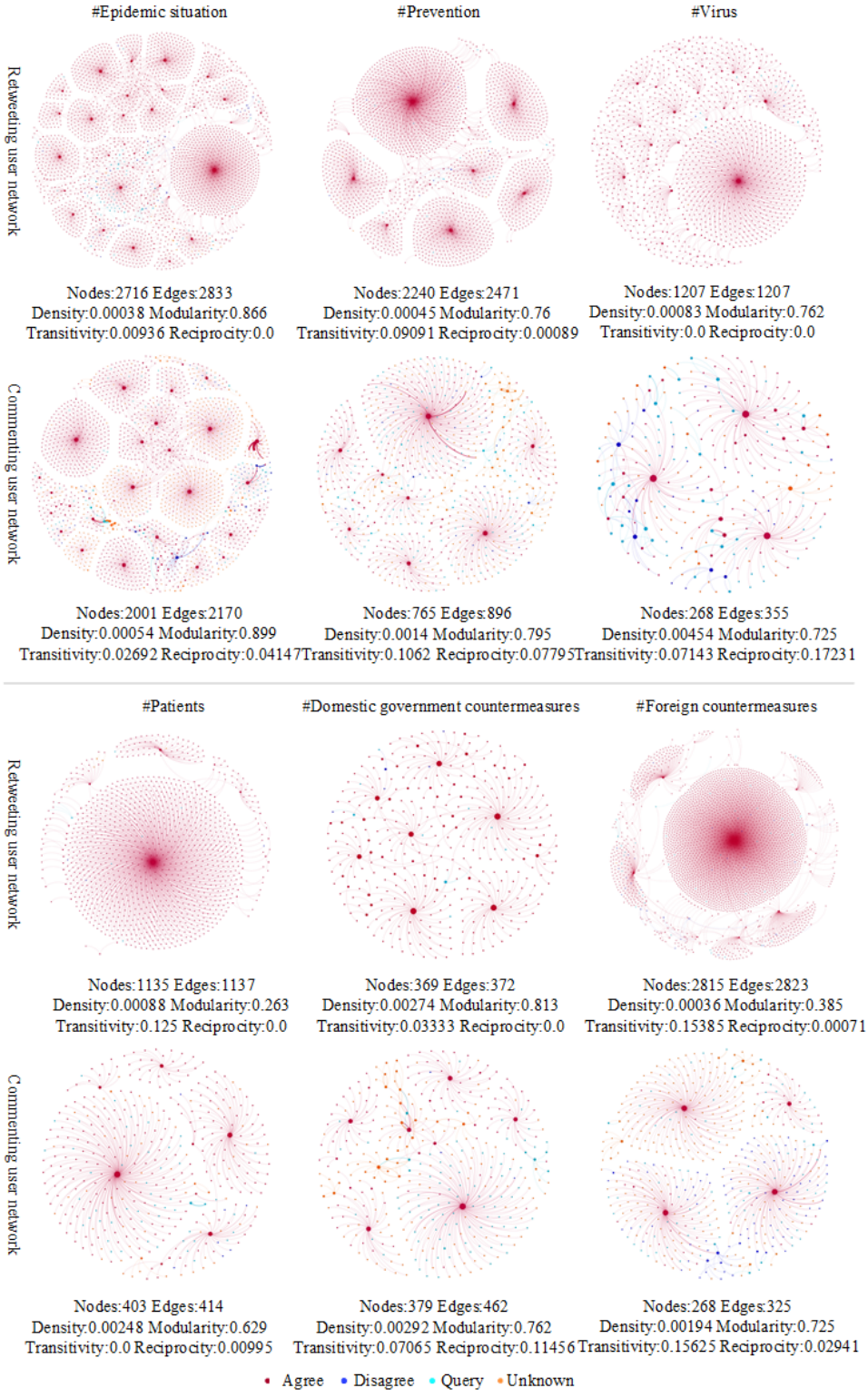


Echo Chamber Effect in Networks of Rumor Rebuttal Under Different Topics

Echo Chamber Effect in User-Level Interaction Networks

In Figure 4, retweeting user networks and commenting user networks are colored by users' attitudes toward rumor rebuttal under the top six topics and visualized using Gephi's Fruchterman Reingold layout algorithm. Each node's size was proportional to its weighted in-degree, each line's thickness was proportional to the edge's weight, and each line's color was consistent with the target node. Compared to retweeting networks, commenting networks were smaller but denser. Except for *virus* and *domestic government countermeasures*, the modularity values of commenting user networks under other topics were higher than those of retweeting user networks. Apart from *patients*, the transitivity and reciprocity—when the first individual chooses the second individual, the second individual also chooses the first individual—of commenting user networks under other topics were higher than those of retweeting user networks. Consequently, compared to retweeting users, commenting users created a more cohesive community with the help of the commenting mechanism [75], and the relationship between users was close and relatively stable [76].

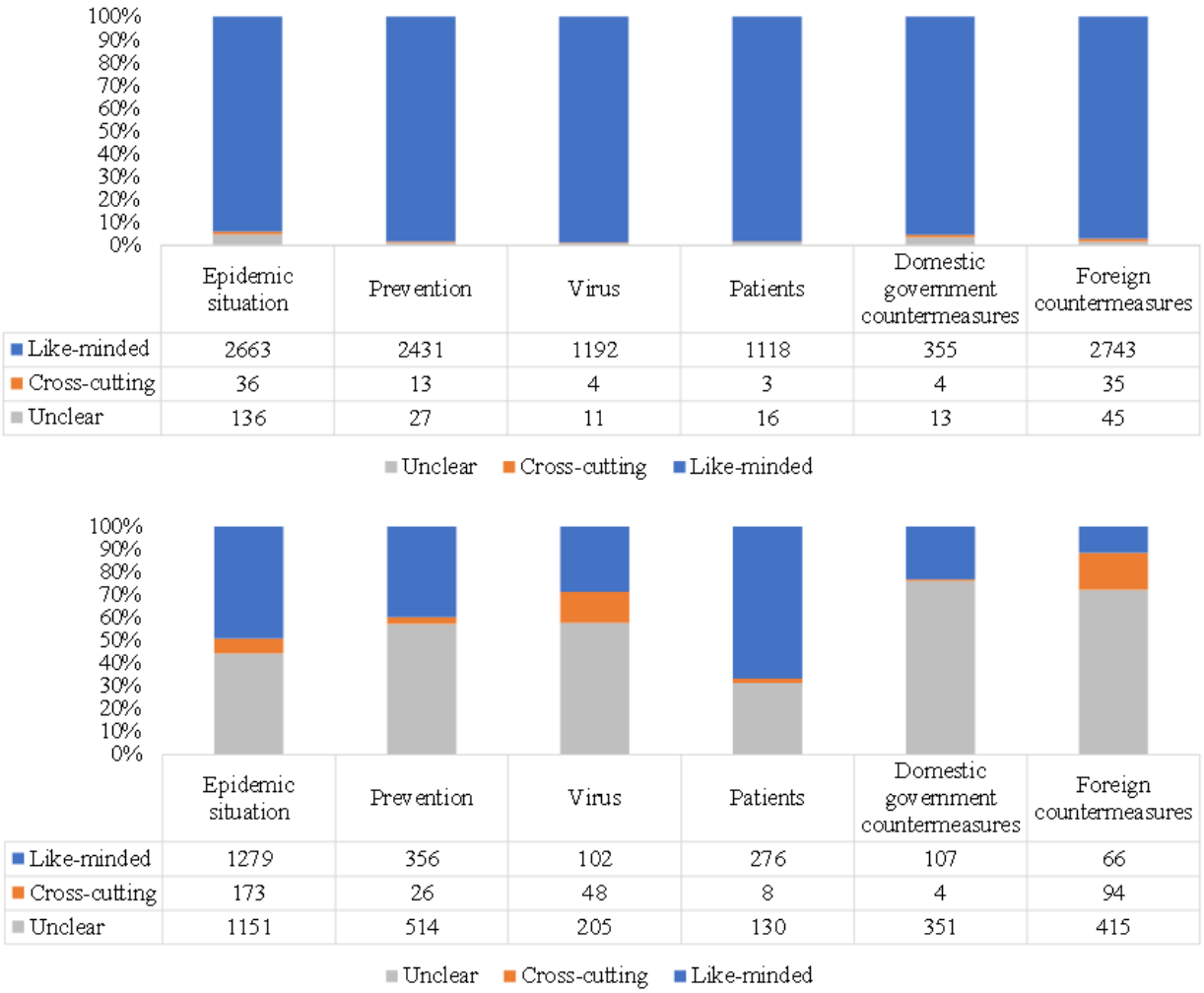
Figure 4. Retweeting user networks and commenting user networks of rumor rebuttal under the top six topics.



As shown in Figure 4, the retweeting and commenting user networks of rumor rebuttal under different topics showed highly modular structures; however, the large clusters in retweeting user networks showed high homophily, while the large clusters in commenting user networks had mixed attitudes. In addition, in retweeting user networks, the attitude of the user node located in the center was mostly *agree*, but in commenting user networks, it might be *agree*, *disagree*, or *query*. In Figure 5, *like-minded* refers to the sum of the edges' weights in which the attitude of the nodes at both ends was either *agree* or *disagree*, *cross-cutting* refers to the sum of the edges' weights in which the attitude of the nodes at one end was *agree* and at the other end was *disagree*, and *unclear* refers to the sum of the edges' weights in which there was at least one node at both ends holding the *query* or *unknown* attitude. In retweeting user networks under different topics, the proportion of interactions between users with the same attitude ranged from 93.9% (2663/2835) to 98.8% (1192/1207), the proportion of

interactions between users with opposite attitudes ranged from 0.3% (3/1137) to 1.3% (36/2835), and the proportion of interactions between users whose attitudes were not clear ranged from 0.9% (11/1207) to 4.8% (138/2835). In commenting user networks, the situation was quite different, with the like-minded interactions accounting for 11.5% (66/575) to 66.7% (276/414), cross-cutting interactions accounting for 0.9% (4/462) to 16.3% (94/575), and unclear interactions accounting for 31.4% (130/414) to 76.0% (351/462). These quantitative indicators revealed the significance of ideology echo chambers in retweeting user networks and the low homophily in commenting user networks.

Figure 5. Homophily based on the users' attitudes in retweeting and commenting user networks toward rumor rebuttal under the top six topics.



	Epidemic situation	Prevention	Virus	Patients	Domestic government countermeasures	Foreign countermeasures
Like-minded	2663	2431	1192	1118	355	2743
Cross-cutting	36	13	4	3	4	35
Unclear	136	27	11	16	13	45

	Epidemic situation	Prevention	Virus	Patients	Domestic government countermeasures	Foreign countermeasures
Like-minded	1279	356	102	276	107	66
Cross-cutting	173	26	48	8	4	94
Unclear	1151	514	205	130	351	415

Echo Chamber Effect in Community-Level Interaction Networks

Figure 6 indicates that the intracommunity heterogeneity value of the retweeting network under each topic was low. Except for *patients* and *domestic government countermeasures*, the intracommunity heterogeneity values of commenting networks under other topics were high. In addition, the intracommunity heterogeneity values of retweeting networks were generally lower than those of commenting networks.

In Figure 7, the retweeting and commenting community networks are colored by communities' attitude scores toward rumor rebuttal under the top six topics; a score of 1 meant that 100% of users in the community held the *agree* attitude and a score of –1 meant that 100% of users in the community held the *disagree* attitude. These were visualized using Gephi's Force Atlas layout algorithm. Each node's size was proportional to the number of users within the community, each line's thickness was proportional to the edge's weight, and each line's color was consistent with the target node. Compared to retweeting networks, the commenting community networks were smaller but denser and had larger nodes. Apart from *patients*, commenting community networks had higher transitivity; except for *patients* and *prevention*, they had higher reciprocity. This meant that, compared to retweeting networks, the interactions between communities in commenting networks were more common and the connections between communities were closer and more stable.

Figure 6. Heterogeneity based on the distribution of users' attitudes within each community node in retweeting (left) and commenting (right) community networks toward rumor rebuttal under the top six topics. The × symbols and circles represent means and outliers of the heterogeneity values, respectively.

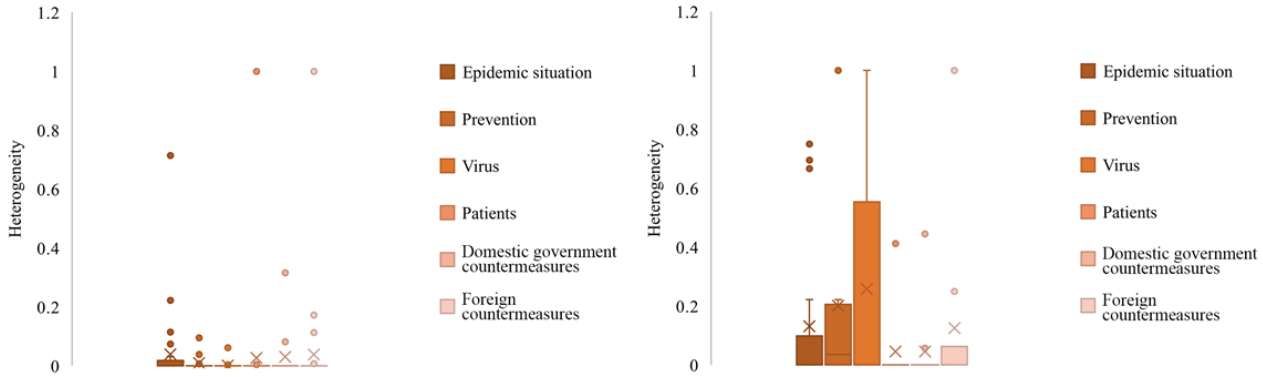
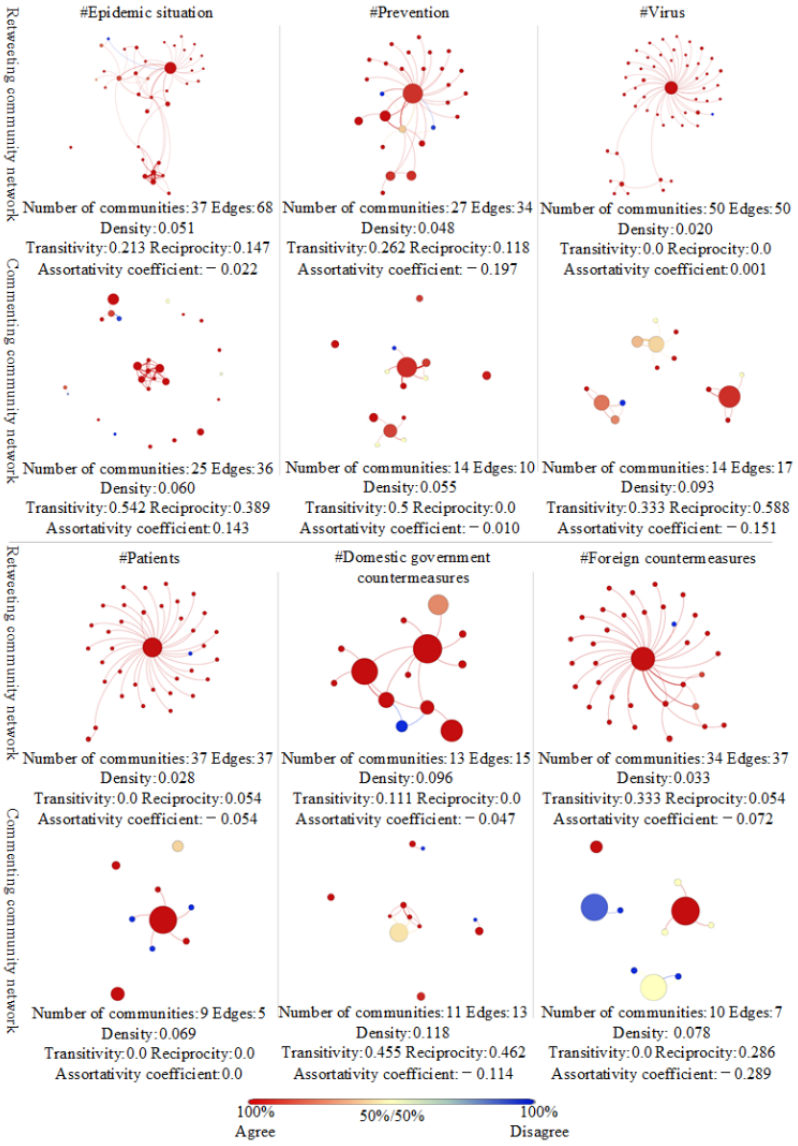


Figure 7. Retweeting community networks and commenting community networks.



It was visually apparent that most communities in retweeting networks had strongly homogeneous attitude distributions toward rumor rebuttal, dominated by the *agree* attitude, while very few communities had a mix of opposite views. However, commenting networks were less segregated and showed more frequent occurrences of communities containing both *agree* and *disagree* attitudes. The assortativity coefficients of retweeting community networks ranged from –0.197 to 0.001, among which *virus* was the highest. The assortativity coefficients of commenting community networks ranged from –0.289 to 0.143, among which *epidemic situation* was the highest. The assortativity coefficients of the retweeting community networks

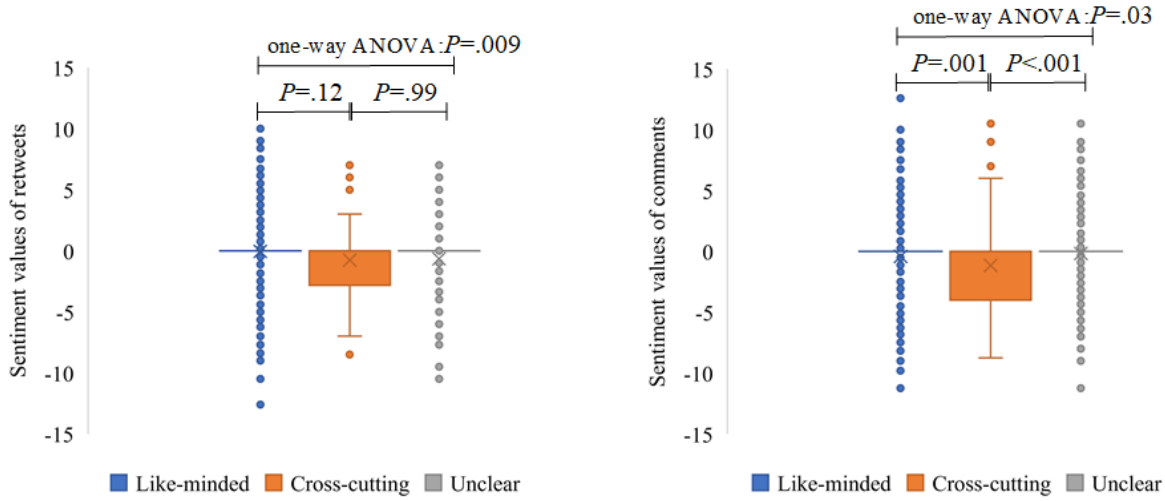
under the topics of *virus*, *domestic government countermeasures*, and *foreign countermeasures* were higher than those of the commenting community networks. The assortativity coefficients of the commenting community networks under the topics of *epidemic situation*, *prevention*, and *patients* were higher than those of the retweeting community networks. All in all, on the one hand, the attitude distribution within communities of retweeting networks under all the topics and commenting networks under *patients* and *domestic government countermeasures* was homogeneous. On the other hand, the interaction homophily between communities in retweeting and commenting networks was low.

The Impact of the Echo Chamber Effect on Information Characteristics of User Interaction Content

The Difference in Information Characteristics Between Interaction Types

As shown in Figure 8, the retweets and comments were generally negative, and the negative values in comments were higher. The average of the sentiment values in cross-cutting interactions was lower than that of the other two kinds of interactions. In addition, it is noteworthy that cross-cutting interactions of retweets and comments contained significantly higher negative values than did like-minded interactions.

Figure 8. The distribution of sentiment values across different kinds of interactions in retweets (left) and comments (right). The × symbols and circles represent means and outliers of the sentiment values, respectively. ANOVA: analysis of variance.



Figures 9 and 10 clearly state that about 4.0% (429/10,845) of the retweets contained information-seeking content and 2.1% (231/10,845) contained information-sharing content, while about 21.9% (1162/5305) of the comments contained information-seeking content and 10.9% (579/5305) contained information-sharing content. Only 2.2% (242/10,845) of the retweets contained uncivil expressions, while 15.9% (845/5305) of the comments contained uncivil expressions. The results of chi-square tests for both retweets and comments (P<.001) claimed that the proportion of information seeking, information sharing, and incivility in cross-cutting interactions was significantly higher than in like-minded interactions, at the significance level of .001.

Figure 9. The proportion of information-seeking, information-sharing, and no information content in different kinds of interactions in retweets (left) and comments (right).

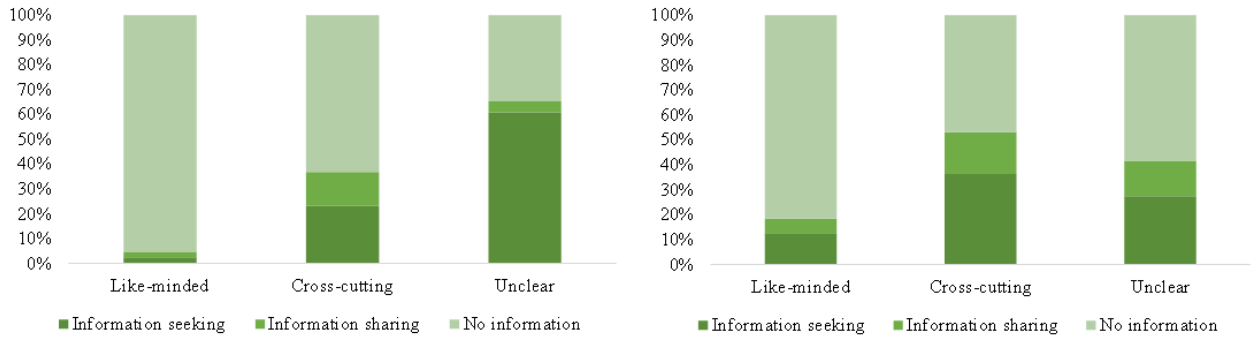
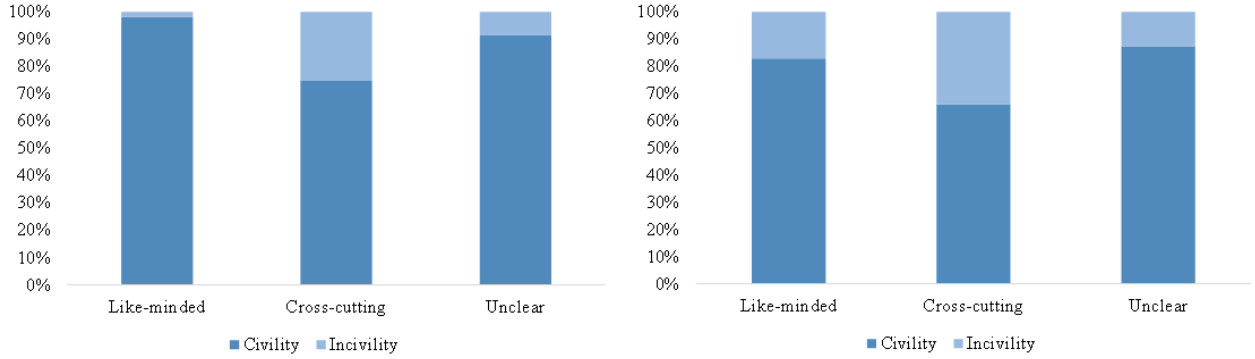


Figure 10. The proportion of civility and incivility in different kinds of interactions in retweets (left) and comments (right).



Correlation of Different Information Characteristics

Figure 11 shows that the sentiment values of retweets or comments containing information-seeking and information-sharing content were significantly different. The average of the sentiment values of retweets without information-seeking and information-sharing content was the highest, and the average sentiment value containing information-sharing content was significantly lower than the one containing information-seeking content. A similar situation occurred with comments. Figure 12 shows that uncivil retweets or comments had higher negative sentiment values than civil retweets or comments.

Figure 11. The distribution of sentiment values in retweets (left) and comments (right) containing information-seeking, information-sharing, and no information content. The × symbols and circles represent means and outliers of the sentiment values, respectively. ANOVA: analysis of variance.

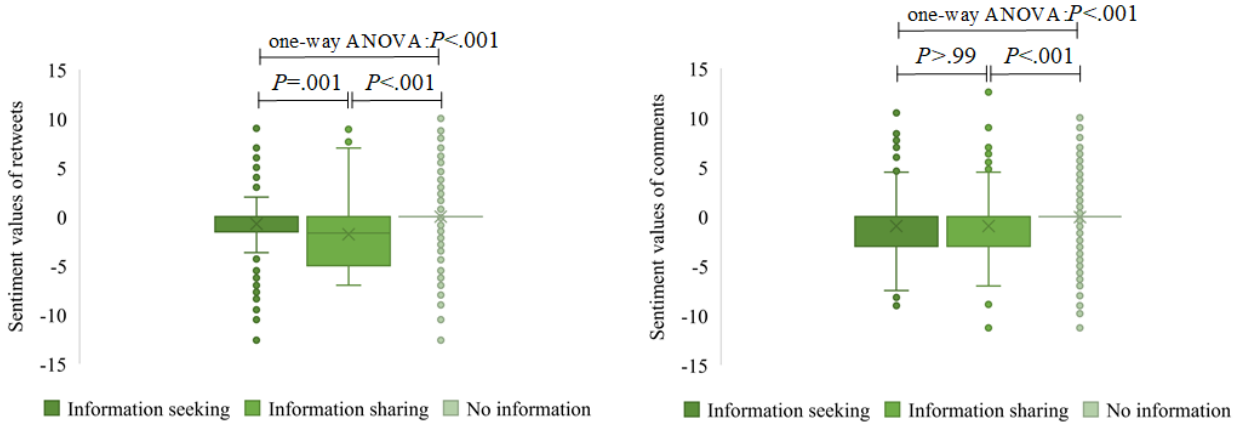


Figure 12. The distribution of sentiment values for retweets (left) and comments (right) containing civil and uncivil content. The × symbols and circles represent means and outliers of the sentiment values, respectively.

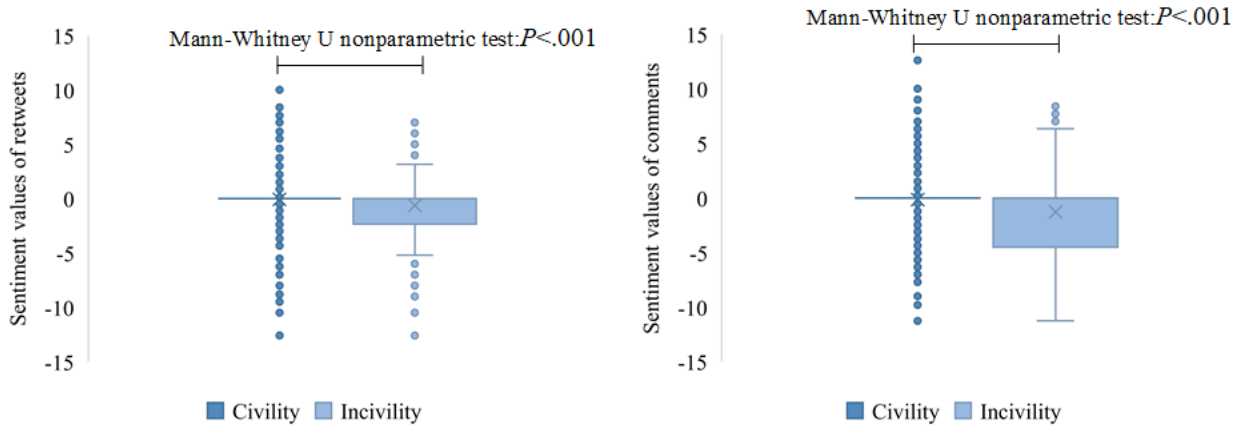
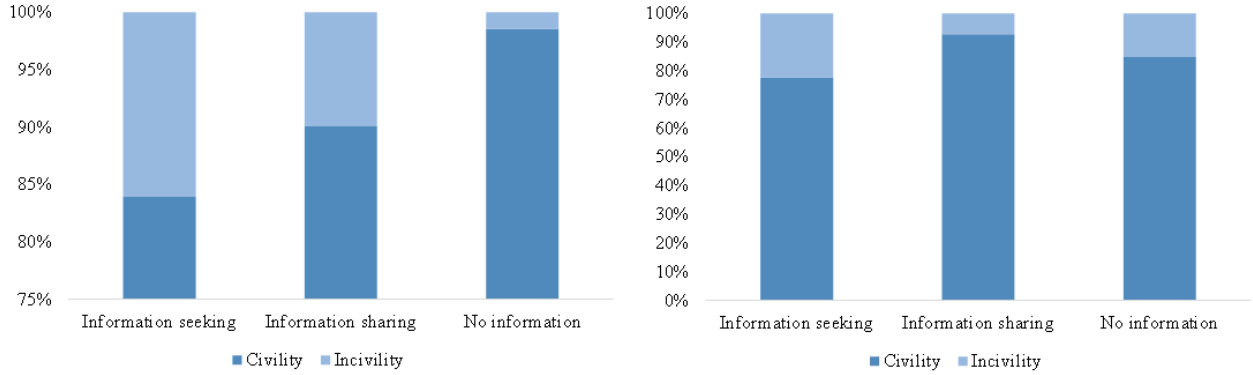


Figure 13 indicates that the proportion of incivility in tweets containing information-seeking content was highest, whether it was in retweets or comments. The results of chi-square tests (P<.001) illustrated that there were significant differences in the proportion of civility and incivility in comments containing information-seeking and information-sharing content, but there were no significant differences in retweets.

Figure 13. The proportion of civility and incivility in retweets (left) and comments (right) containing information-seeking, information-sharing, and no information content.



Discussion

Principal Findings

Considering the important role of the dissemination of rumor rebuttal on social media regarding rumor control and disease containment during public health crises, in this research, we used content analysis, sentiment analysis, social network analysis, and statistical analysis to roughly evaluate the effectiveness of rumor rebuttal, analyze the echo chamber effect based on attitudes in users' retweeting and commenting behaviors toward rumor rebuttal under different topics, and analyze the impact of the echo chamber effect on information characteristics of user interaction content. Firstly, we analyzed the distribution of attitudes and whether there was an echo chamber effect in the attitude choice of individuals when retweeting or commenting on others' tweets. Secondly, we tested the heterogeneity of attitude distribution within communities and the homophily of interactions between communities. Based on the results at user and community levels, we made a comprehensive judgment. Finally, we examined the content of user interaction from three dimensions of sentiment expression, information seeking and sharing, and civility to test the impact of the echo chamber effect.

The Echo Chamber Effect in Rumor Rebuttal Communication

Retweeting indicates a desire to increase the visibility of a given message; comments are a way of online collective debate around the topic of a tweet [77]. Therefore, comments are more likely to include diversified pros and cons–related feedback on rumor rebuttal under different topics [16]. Moreover, social media users who participate in different topics of rumor rebuttal experience different degrees of the echo chamber effect in attitude selection based on different interaction mechanisms. The user-level interaction networks indicated a high frequency of like-minded interactions and a low frequency of cross-cutting interactions in retweeting networks of all topics; however, they indicated a low frequency of like-minded interactions and a high frequency of interactions containing users without clear attitudes in commenting networks of all topics. Consistent with the research of Wang and Song [15], Tsai et al [40], and Williams et al [41], our results once again emphasized the outstanding performance of attitude-based echo chambers in retweeting networks and the frequent interaction of dissimilar opinions in commenting networks. Retweeting often implies endorsement of either the individual tweet or its original author [41]. Thus, users tend to retweet others who have attitudes consistent with their own, which is in line with the goal of an individual developing social identity, while commenting usually serves as an open channel for viewpoint collision and fusion. However, our results might conflict with the research of Shin et al [39] and Zollo et al [16]. Results from the former study might be due to the influence of political context on selective exposure. Results from the latter study might lie in platform differences. As Facebook focuses on reciprocal social interaction and Weibo, like Twitter, concentrates more on the sharing of opinions with the reduction of social pressure brought about by anonymity, echo chambers exist in users' commenting networks on Facebook and not on Weibo [78].

The community-level interaction networks strongly suggested that homophily occurred in attitude distribution within communities in retweeting networks under all topics and in commenting networks under *patients* and *domestic government countermeasures* topics. Like the conclusions from Williams et al [41] on the controversial topics about climate change, most individuals who engaged in online discussions were embedded within communities of like-minded users; such self-reinforcing *echo chambers* could prevent engagement with alternative attitudes and promote extreme views [79]. We also identified mixed-attitude communities in commenting networks corresponding to certain topics, in which users were frequently exposed to a diversity of attitudes. Williams et al [41] characterized such communities as *open forums*. Nevertheless, there was not a clear tendency for a whole community to interact most strongly with other communities of similar attitude composition either in retweeting or commenting networks. This indicated that the echo chamber effect was not significant at the level of community interaction.

The Impact of the Echo Chamber Effect on Interaction Content

Online content of user interactions using different mechanisms had different properties. This reflected different motivations for using social media. Compared with retweeting, commenting was used more for sentiment expression, information seeking, and information sharing; it also contained more uncivil terms. These might be attributed to the comment mechanism acting as

an *open forum* to reduce the individual's exposure ratio in social networks [80]. Online social networks are recognized sites of both the construction of social identities [81,82] and their linguistic performance [83]. Social identity theory [81] asserts that willingness to negatively engage with out-group members is a way of affirming membership of the in-group. In other words, interactions with alternative attitudes are often accompanied by greater negative sentiment. In addition, to defend collective sensemaking, it is inevitable to breed uncivil terms. It is worth noting that most like-minded interactions did not seek or share information; on the contrary, cross-cutting interactions did seek or share information. This suggested that the echo chamber effect dealt with the potential adverse effects of knowledge flow and group wisdom gathering [15]. The research also showed that online users' information-seeking behavior was accompanied by incivility, and information-sharing behavior was accompanied by a more negative sentiment, which was often accompanied by incivility. This was very detrimental to any kind of meaningful interaction between users.

Theoretical Contributions and Practical Implications

There are three main contributions. Firstly, previous research studies on the effectiveness of rumor rebuttal were mostly based on self-reported perception and attitude data, and they ignored or simplified the actual network structure of rumor rebuttal dissemination and users' cognition as well as their decision-making and interaction behaviors. This research analyzed the echo chamber effect in users' responses to rumor rebuttal under different topics, based on naturally occurring online behavior data. This attempt filled the gap between rumor rebuttal and echo chamber research and provided a new research area. Secondly, when analyzing the echo chamber effect, this study started from multiple dimensions, such as topic, interaction mechanism, and interaction level, and used visualization combined with qualitative and quantitative indicators to enrich and enhance the robustness of the conclusions. Thirdly, this study not only determined the existence and degree of the echo chamber effect but also analyzed its impact on the characteristics of interactive content (ie, sentiment, information, and civility), introducing sentiment intensity to improve the accuracy of the analysis.

In the process of rumor rebuttal dissemination under different topics, the distribution of attitudes by users who retweeted or commented was different, and the significance of an attitude-based echo chamber effect was also different. Social media platform managers should systematically monitor users' attitudes and should achieve multidimensional governance by topic, network (ie, retweeting and commenting), and community (ie, within the network community and between the network communities) in order to prevent or alleviate group polarization. Notably, managers should strengthen the sentiment guidance for users who denied the rumor rebuttal in the open commenting forum to guard against sentiment infection. Simultaneously, the censorship of uncivil tweets should be increased to update the network environment, so that information seeking and sharing can become more efficient.

Limitations

It should be noted that Weibo's @ function is often used to initiate conversations with target users. This kind of targeted request, different from nontargeted exposure of retweeting and commenting, may exhibit different degrees of the echo chamber effect and then affect the characteristics of interactive content. Additional research should be conducted in the future. This research showed that users' sentiments were negative in retweets and comments. It is meaningful to explore how the echo chamber effect affects different types of negative sentiments, such as anger and fear.

Conclusions

This study first analyzed the distribution of attitudes by users who retweeted or commented on rumor rebuttal on Weibo in the early stage of the COVID-19 epidemic and found that the effectiveness of rumor rebuttal varied among different topics. The study then dug deeply into the existence and degree of an attitude-based echo chamber effect and its impact on causing sentiment resonance, motivating information seeking and sharing, and breeding uncivil speech. The findings confirmed that retweeting played an essential role in promoting polarization, and commenting played a role in consensus building. The findings showed that there might not be a significant echo chamber effect in community interactions and verified that the echo chamber effect did have an essential impact on interactive content (ie, sentiment, information, and civility). Specifically, polarization caused cross-cutting interactions to contain more negative sentiment, which was associated with incivility, and caused like-minded interactions to contain less meaningful information seeking and sharing.

Acknowledgments

Funding was provided by the National Natural Science Foundation of China (grants 71661167007 and 71420107026) and by the National Key Research and Development Program of China (grant 2018YFC0806904-03).

Conflicts of Interest

None declared.

References

1. Thelwall M, Thelwall S. A thematic analysis of highly retweeted early COVID-19 tweets: Consensus, information, dissent and lockdown life. Aslib J Inf Manage 2020 Oct 23;72(6):945-962. [doi: 10.1108/ajim-05-2020-0134]

2. Tian H, Liu Y, Li Y, Wu C, Chen B, Kraemer M, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science 2020 May 08;368(6491):638-642 [FREE Full text] [doi: 10.1126/science.abb6105] [Medline: 32234804]
3. Zubiaga A, Liakata M, Procter R, Wong Sak Hoi G, Tolmie P. Analysing how people orient to and spread rumours in social media by looking at conversational threads. PLoS One 2016 Mar 4;11(3):e0150989 [FREE Full text] [doi: 10.1371/journal.pone.0150989] [Medline: 26943909]
4. The Lancet. COVID-19: Fighting panic with information. Lancet 2020 Feb 22;395(10224):537 [FREE Full text] [doi: 10.1016/S0140-6736(20)30379-2] [Medline: 32087777]
5. Larson HJ. Blocking information on COVID-19 can fuel the spread of misinformation. Nature 2020 Apr;580(7803):306. [doi: 10.1038/d41586-020-00920-w] [Medline: 32231320]
6. Lipsitch M, Swerdlow DL, Finelli L. Defining the epidemiology of Covid-19 — Studies needed. N Engl J Med 2020 Mar 26;382(13):1194-1196. [doi: 10.1056/nejmp2002125]
7. Pal A, Banerjee S. Internet users beware, you follow online health rumors (more than counter-rumors) irrespective of risk propensity and prior endorsement. Inf Technol People 2020 Dec 01:1. [doi: 10.1108/ITP-02-2019-0097]
8. Cao Y, An B. Weibo gives rumors no time to thrive. China Daily. 2011 Nov 04. URL: https://www.chinadaily.com.cn/china/2011-11/04/content_14035335.htm [accessed 2021-03-22]
9. Zeng J, Chan CH, Fu KW. How social media construct “truth” around crisis events: Weibo's rumor management strategies after the 2015 Tianjin blasts. Policy Internet 2017;9(3):297-320. [doi: 10.1002/poi3.155]
10. Zeng J, Burgess J, Bruns A. Is citizen journalism better than professional journalism for fact-checking rumours in China? How Weibo users verified information following the 2015 Tianjin blasts. Glob Media China 2019 Mar 25;4(1):13-35 [FREE Full text] [doi: 10.1177/2059436419834124]
11. Petrova P, Cialdini R. Fluency of consumption imagery and the backfire effects of imagery appeals. J Consum Res 2005 Dec;32(3):442-452. [doi: 10.1086/497556]
12. Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, et al. The spreading of misinformation online. Proc Natl Acad Sci U S A 2016 Jan 19;113(3):554-559 [FREE Full text] [doi: 10.1073/pnas.1517441113] [Medline: 26729863]
13. Dubois E, Blank G. The echo chamber is overstated: The moderating effect of political interest and diverse media. Inf Commun Soc 2018 Jan 29;21(5):729-745 [FREE Full text] [doi: 10.1080/1369118x.2018.1428656]
14. Li W, Peng J. Simulation experiment on echo chamber effect of social networks communication. Mod Commun 2019;41(4):142-149.
15. Wang X, Song Y. Viral misinformation and echo chambers: The diffusion of rumors about genetically modified organisms on social media. Internet Res 2020 Jun 22;30(5):1547-1564. [doi: 10.1108/intr-11-2019-0491]
16. Zollo F, Bessi A, Del Vicario M, Scala A, Caldarelli G, Shekhtman L, et al. Debunking in a world of tribes. PLoS One 2017;12(7):e0181821 [FREE Full text] [doi: 10.1371/journal.pone.0181821] [Medline: 28742163]
17. Garrett RK. The “echo chamber” distraction: Disinformation campaigns are the problem, not audience fragmentation. J Appl Res Mem Cogn 2017 Dec;6(4):370-376. [doi: 10.1016/j.jarmac.2017.09.011]
18. DiFonzo N, Bordia P. Rumor Psychology: Social and Organizational Approaches. Washington, DC: American Psychological Association; 2007.
19. Jung C. Contribution to the psychology of rumour. Indiana Assoc Health Phys Educ Recreat Dance J 1909;18(1):1-26.
20. Ozturk P, Li H, Sakamoto Y. Combating rumor spread on social media: The effectiveness of refutation and warning. In: Proceedings of the 48th Hawaii International Conference on System Sciences.: IEEE; 2015 Mar 30 Presented at: 48th Hawaii International Conference on System Sciences; January 5-8, 2015; Kauai, HI p. 2406-2414. [doi: 10.1109/hicss.2015.288]
21. Tanaka Y, Sakamoto Y, Matsuka T. Toward a social-technological system that inactivates false rumors through the critical thinking of crowds. In: Proceedings of the 46th Hawaii International Conference on System Sciences.: IEEE; 2013 Mar 18 Presented at: 46th Hawaii International Conference on System Sciences; January 7-10, 2013; Wailea, HI p. 649-658. [doi: 10.1109/hicss.2013.557]
22. Pal A, Chua AY, Goh DH. Does KFC sell rat? Analysis of tweets in the wake of a rumor outbreak. Aslib J Inf Manage 2017 Nov 20;69(6):660-673. [doi: 10.1108/ajim-01-2017-0026]
23. Pal A, Chua AY, Hoe-Lian Goh D. Debunking rumors on social media: The use of denials. Comput Human Behav 2019 Jul;96:110-122. [doi: 10.1016/j.chb.2019.02.022]
24. Berinsky AJ. Rumors, truths, and reality: A study of political misinformation. In: Proceedings of the 69th Annual National Conference of the Midwest Political Science Association. 2011 Presented at: 69th Annual National Conference of the Midwest Political Science Association; March 31-April 3, 2011; Chicago, IL URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.945.4298&rep=rep1&type=pdf
25. Bordia P, DiFonzo N, Haines R, Chaseling E. Rumors denials as persuasive messages: Effects of personal relevance, source, and message characteristics. J Appl Soc Psychol 2005 Jun;35(6):1301-1331. [doi: 10.1111/j.1559-1816.2005.tb02172.x]
26. Bordia P, DiFonzo N, Travers V. Denying rumors of organizational change: A higher source is not always better. Commun Res Rep 1998 Mar;15(2):188-197. [doi: 10.1080/08824099809362113]

27. DiFonzo N, Bordia P. How top PR professionals handle hearsay: Corporate rumors, their effects, and strategies to manage them. Public Relat Rev 2000 Jun;26(2):173-190. [doi: 10.1016/s0363-8111(00)00039-4]
28. Chen J, Liu Y, Deng S. Research on user reviews of government rumor-refuting information and factors influencing their emotional tendencies. Inf Sci China 2017;35(12):72-75.
29. Ruan W, Xia Z. Analysis of influencing factors of social media users? Willingness to share counter-rumors. Sci Manage China 2020;40(2):39-44.
30. Chen Y. Study on the spread and control of internet rumors of public health emergencies: Text analysis of internet rumors based on COVID-19 epidemic. E-Government 2020;17(6):2-11. [doi: 10.16582/j.cnki.dzzw.2020.06.001]
31. Huang J. Rumor rebuttal practice of government WeChat during COVID-19 epidemic: A case study of "Ruian release". J News Res China 2020 May;11(10):133-134.
32. Li Z, Zhang Q, Wang Y, Wang S. Social media rumor refuter feature analysis and crowd identification based on XGBoost and NLP. Appl Sci (Basel) 2020 Jul 08;10(14):4711 [FREE Full text] [doi: 10.3390/app10144711]
33. Anthony S. The influence of personal characteristics on rumor knowledge and transmission among the deaf. Am Ann Deaf 1992 Mar;137(1):44-47. [doi: 10.1353/aad.2012.0413] [Medline: 1605100]
34. Einwiller S, Kamins M. Rumor has it: The moderating effect of identification on rumor impact and the effectiveness of rumor refutation. J Appl Soc Psychol 2008;38(9):2248-2272. [doi: 10.1111/j.1559-1816.2008.00390.x]
35. Mael F, Ashforth BE. Alumni and their alma mater: A partial test of the reformulated model of organizational identification. J Organ Behav 1992 Mar;13(2):103-123. [doi: 10.1002/job.4030130202]
36. DiFonzo N. Conspiracy rumor psychology. In: Uscinski JE, editor. Conspiracy Theories and the People Who Believe Them. Oxford, UK: Oxford University Press; 2018:257-268.
37. Bruns A. Echo chamber? What echo chamber? Reviewing the evidence. In: Proceedings of the 6th Biennial Future of Journalism Conference (FOJ17). 2017 Presented at: 6th Biennial Future of Journalism Conference (FOJ17); September 14-15, 2017; Cardiff, UK p. 1 URL: http://snurb.info/files/2017/Echo%20Chamber.pdf
38. Himelboim I, McCreery S, Smith M. Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter. J Comput Mediat Commun 2013 Jan 25;18(2):40-60. [doi: 10.1111/jcc4.12001]
39. Shin J, Jian L, Driscoll K, Bar F. Political rumoring on Twitter during the 2012 US presidential election: Rumor diffusion and correction. New Media Soc 2016 Mar 08;19(8):1214-1235. [doi: 10.1177/1461444816634054]
40. Tsai WS, Tao W, Chuan C, Hong C. Echo chambers and social mediators in public advocacy issue networks. Public Relat Rev 2020 Mar 01;46(1):101882. [doi: 10.1016/j.pubrev.2020.101882]
41. Williams HTP, McMurray JR, Kurz T, Lambert FH. Network analysis reveals open forums and echo chambers in social media discussions of climate change. Glob Environ Change 2015 May;32:126-138. [doi: 10.1016/j.gloenvcha.2015.03.006]
42. Wang B, Zhuang J. Crisis information distribution on Twitter: A content analysis of tweets during Hurricane Sandy. Nat Hazards 2017 Jun 20;89(1):161-181. [doi: 10.1007/s11069-017-2960-x]
43. David CC, Ong JC, Legara EFT. Tweeting supertyphoon Haiyan: Evolving functions of Twitter during and after a disaster event. PLoS One 2016;11(3):e0150190 [FREE Full text] [doi: 10.1371/journal.pone.0150190] [Medline: 27019425]
44. Person B, Sy F, Holton K, Govert B, Liang A, National Center for Inectious Diseases/SARS Community Outreach Team. Fear and stigma: The epidemic within the SARS outbreak. Emerg Infect Dis 2004 Feb;10(2):358-363 [FREE Full text] [doi: 10.3201/eid1002.030750] [Medline: 15030713]
45. Yang JZ, Chu H. Who is afraid of the Ebola outbreak? The influence of discrete emotions on risk perception. J Risk Res 2016 Oct 28;21(7):834-853. [doi: 10.1080/13669877.2016.1247378]
46. Dong W, Tao J, Xia X, Ye L, Xu H, Jiang P, et al. Public emotions and rumors spread during the COVID-19 epidemic in China: Web-based correlation study. J Med Internet Res 2020 Nov 25;22(11):e21933 [FREE Full text] [doi: 10.2196/21933] [Medline: 33112757]
47. Zeng R, Zhu D. A model and simulation of the emotional contagion of netizens in the process of rumor refutation. Sci Rep 2019 Oct 02;9(1):14164 [FREE Full text] [doi: 10.1038/s41598-019-50770-4] [Medline: 31578405]
48. Einwiller SA. Enough is enough! When identification no longer prevents negative corporate associations. J Acad Mark Sci 2006 Apr 01;34(2):185-194. [doi: 10.1177/0092070305284983]
49. Wollebæk D, Karlsen R, Steen-Johnsen K, Enjolras B. Anger, fear, and echo chambers: The emotional basis for online behavior. Soc Media Soc 2019 Apr 09;5(2):1-14 [FREE Full text] [doi: 10.1177/2056305119829859]
50. Xu W, Zhang C. Sentiment, richness, authority, and relevance model of information sharing during social crises—The case of #MH370 tweets. Comput Human Behav 2018 Dec;89:199-206. [doi: 10.1016/j.chb.2018.07.041]
51. Chew C, Eysenbach G. Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
52. Merrifield N, Palenchar M. Uncertainty reduction strategies via Twitter: The 2011 wildfire threat to Los Alamos National Laboratory. In: Proceedings of the Association for Education in Journalism and Mass Communication (AEJMC) Annual Conference. 2012 Presented at: Association for Education in Journalism and Mass Communication (AEJMC) Annual Conference; August 9-12, 2012; Chicago, IL p. 1.
53. Hughes AL, Palen L. Twitter adoption and use in mass convergence and emergency events. Int J Emerg Manag 2009;6(3/4):248-260. [doi: 10.1504/ijem.2009.031564]

54. Rowe I. Civility 2.0: A comparative analysis of incivility in online political discussion. Inf Commun Soc 2014 Jul 22;18(2):121-138. [doi: 10.1080/1369118x.2014.940365]
55. Berg J. The impact of anonymity and issue controversiality on the quality of online discussion. J Inf Technol Polit 2016 Mar 07;13(1):37-51. [doi: 10.1080/19331681.2015.1131654]
56. Chen L, Wu X, Li M. Formation and fragmentation within a networked public sphere: Social media debates on Traditional Chinese Medicine. Telematics Inform 2018 Dec;35(8):2219-2231. [doi: 10.1016/j.tele.2018.08.008]
57. Berger J. Arousal increases social transmission of information. Psychol Sci 2011 Jul;22(7):891-893. [doi: 10.1177/0956797611413294] [Medline: 21690315]
58. Hasell A, Weeks BE. Partisan provocation: The role of partisan news use and emotional responses in political information sharing in social media. Hum Commun Res 2016 Jul 08;42(4):641-661. [doi: 10.1111/hcre.12092]
59. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020 Feb;395(10223):470-473. [doi: 10.1016/s0140-6736(20)30185-9]
60. Yang J, Lee S. Framing the MERS information crisis: An analysis on online news media's rumour coverage. J Conting Crisis Manage 2020 Apr 27;28(4):386-398. [doi: 10.1111/1468-5973.12292]
61. Gu H, Chen B, Zhu H, Jiang T, Wang X, Chen L, et al. Importance of internet surveillance in public health emergency control and prevention: Evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks. J Med Internet Res 2014 Jan 17;16(1):e20 [FREE Full text] [doi: 10.2196/jmir.2911] [Medline: 24440770]
62. Ma J, Luo Y. The classification of rumour standpoints in online social network based on combinatorial classifiers. J Inf Sci 2019 Feb 21;46(2):191-204. [doi: 10.1177/0165551519828619]
63. Krippendorff K. Computing Krippendorff's alpha-reliability. University of Pennsylvania ScholarlyCommons. 2011 Jan 25. URL: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1043&context=asc_papers [accessed 2021-03-22]
64. Li Z, Zhang Q, Du X, Ma Y, Wang S. Social media rumor refutation effectiveness: Evaluation, modelling and enhancement. Inf Process Manag 2021 Jan;58(1):102420. [doi: 10.1016/j.ipm.2020.102420]
65. An L, Ou M. Social network sentiment map of the stakeholders in public health emergencies. Libr Inf Serv China 2017 Oct;61(20):120-130.
66. Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using NetworkX. In: Proceedings of the 7th Python in Science Conference (SciPy2008). 2008 Presented at: 7th Python in Science Conference (SciPy2008); August 19-24, 2008; Pasadena, CA p. 11-15 URL: http://conference.scipy.org/proceedings/scipy2008/paper_2/full_text.pdf
67. Grandjean M. GEPHI – Introduction to network analysis and visualization. Serval, Université de Lausanne. 2015. URL: https://serval.unil.ch/resource/serval:BIB_0F0EB41780EB.P001/REF.pdf [accessed 2021-03-22]
68. Kauffman J, Kittas A, Bennett L, Tsoka S. DyCoNet: A Gephi plugin for community detection in dynamic complex networks. PLoS One 2014;9(7):e101357 [FREE Full text] [doi: 10.1371/journal.pone.0101357] [Medline: 25000497]
69. Newman MEJ. Assortative mixing in networks. Phys Rev Lett 2002 Oct 28;89(20):208701. [doi: 10.1103/physrevlett.89.208701]
70. Noldus R, Van Mieghem P. Assortativity in complex networks. J Complex Netw 2015 Mar 26;3(4):507-542. [doi: 10.1093/comnet/cnv005]
71. Thedchanamoorthy G, Piraveenan M, Kasthuriratna D, Senanayake U. Node assortativity in complex networks: An alternative approach. Procedia Comput Sci 2014;29:2449-2461. [doi: 10.1016/j.procs.2014.05.229]
72. McKnight PE, Najab J. Mann-Whitney U test. In: Weiner IB, Craighead WE, editors. The Corsini Encyclopedia of Psychology. 4th edition. Hoboken, NJ: John Wiley & Sons; 2010:960.
73. Bewick V, Cheek L, Ball J. Statistics review 9: One-way analysis of variance. Crit Care 2004 Apr;8(2):130-136 [FREE Full text] [doi: 10.1186/cc2836] [Medline: 15025774]
74. Berkson J. Some difficulties of interpretation encountered in the application of the chi-square test. J Am Stat Assoc 1938 Sep;33(203):526-536. [doi: 10.1080/01621459.1938.10502329]
75. Kartun-Giles AP, Bianconi G. Beyond the clustering coefficient: A topological analysis of node neighbourhoods in complex networks. Chaos Solitons Fractals: X 2019 Mar;1:100004. [doi: 10.1016/j.csfx.2019.100004]
76. Weng L, Menczer F, Ahn Y. Virality prediction and community structure in social networks. Sci Rep 2013;3:2522 [FREE Full text] [doi: 10.1038/srep02522] [Medline: 23982106]
77. Ellison N, Steinfield C, Lampe C. The benefits of Facebook "friends": Social capital and college students' use of online social network sites. J Comput Mediat Commun 2007 Jul 01;12(4):1143-1168. [doi: 10.1111/j.1083-6101.2007.00367.x]
78. Hughes DJ, Rowe M, Batey M, Lee A. A tale of two sites: Twitter vs Facebook and the personality predictors of social media usage. Comput Human Behav 2012 Mar;28(2):561-569. [doi: 10.1016/j.chb.2011.11.001]
79. Sunstein CR. Republic.com 2.0. Princeton, NJ: Princeton University Press; 2007.
80. Liu Y, Jin X, Shen H, Cheng X. Do rumors diffuse differently from non-rumors? A systematically empirical analysis in Sina Weibo for rumor identification. In: Proceedings of the 21st Pacific-Asia Conference on Knowledge Discovery and Data Mining. Cham, Switzerland: Springer; 2017 Presented at: 21st Pacific-Asia Conference on Knowledge Discovery and Data Mining; May 23-26, 2017; Jeju, South Korea p. 407-420. [doi: 10.1007/978-3-319-57454-7_32]

81. Tajfel H, Turner J. An integrative theory of intergroup conflict. In: Austin WG, Worchel S, editors. The Social Psychology of Intergroup Relations. Monterey, CA: Brooks Cole Publishing Company; 1979:33-47.
82. Zhao S, Grasmuck S, Martin J. Identity construction on Facebook: Digital empowerment in anchored relationships. Comput Human Behav 2008 Sep;24(5):1816-1836. [doi: 10.1016/j.chb.2008.02.012]
83. Tamburrini N, Cinnirella M, Jansen VA, Bryden J. Twitter users change word usage according to conversation-partner social identity. Soc Networks 2015 Jan;40:84-89. [doi: 10.1016/j.socnet.2014.07.004]

Abbreviations

ANOVA: analysis of variance

Edited by C Basch; submitted 07.01.21; peer-reviewed by S Doan, R Menhas; comments to author 19.02.21; revised version received 03.03.21; accepted 04.03.21; published 25.03.21.

Please cite as:
Wang D, Qian Y
Echo Chamber Effect in Rumor Rebuttal Discussions About COVID-19 in China: Social Media Content and Network Analysis Study
J Med Internet Res 2021;23(3):e27009
URL: https://www.jmir.org/2021/3/e27009
doi:10.2196/27009
PMID:33690145

Original Paper

Impact of a Serious Game (Escape COVID-19) on the Intention to Change COVID-19 Control Practices Among Employees of Long-term Care Facilities: Web-Based Randomized Controlled Trial

Mélanie Suppan[1], MD; Mohamed Abbas[2*], MD; Gaud Catho[2*], MD; Loric Stuby[3*], CAS; Simon Regard[4,5], MD, MPH; Sophia Achab[6,7], MD, PhD; Stephan Harbarth[2], MD, MSc; Laurent Suppan[4], MD

[1]Division of Anesthesiology, Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

[2]Infection Control Programme and WHO Collaborating Centre on Patient Safety, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

[3]Genève TEAM Ambulances, Geneva, Switzerland

[4]Division of Emergency Medicine, Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

[5]Division of the Surgeon General, Geneva Directorate of Health, Geneva, Switzerland

[6]Specialized Facility in Behavioral Addictions ReConnecte, University of Geneva Hospitals, Geneva, Switzerland

[7]WHO Collaborating Center in Training and Research in Mental Health, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[*]these authors contributed equally

Corresponding Author:
Mélanie Suppan, MD
Division of Anesthesiology
Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine
University of Geneva Hospitals and Faculty of Medicine
Rue Gabrielle-Perret-Gentil 4
Geneva, 1211
Switzerland
Phone: 41 223723311
Email: melanie.suppan@hcuge.ch

Abstract

Background: Most residents of long-term care facilities (LTCFs) are at high risk of complications and death following SARS-CoV-2 infection. In these facilities, viral transmission can be facilitated by shortages of human and material resources, which can lead to suboptimal application of infection prevention and control (IPC) procedures. To improve the dissemination of COVID-19 IPC guidelines, we developed a serious game called “Escape COVID-19” using Nicholson’s RECIPE for meaningful gamification, as engaging serious games have the potential to induce behavioral change.

Objective: As the probability of executing an action is strongly linked to the intention of performing it, the objective of this study was to determine whether LTCF employees were willing to change their IPC practices after playing “Escape COVID-19.”

Methods: This was a web-based, triple-blind, randomized controlled trial, which took place between November 5 and December 4, 2020. The health authorities of Geneva, Switzerland, asked the managers of all LTCFs under their jurisdiction to forward information regarding the study to all their employees, regardless of professional status. Participants were unaware that they would be randomly allocated to one of two different study paths upon registration. In the control group, participants filled in a first questionnaire designed to gather demographic data and assess baseline knowledge before accessing regular online IPC guidelines. They then answered a second questionnaire, which assessed their willingness to change their IPC practices and identified the reasons underlying their decision. They were then granted access to the serious game. Conversely, the serious game group played “Escape COVID-19” after answering the first questionnaire but before answering the second one. This group accessed the control material after answering the second set of questions. There was no time limit. The primary outcome was the proportion of LTCF employees willing to change their IPC practices. Secondary outcomes included the factors underlying

participants' decisions, the domains these changes would affect, changes in the use of protective equipment items, and attrition at each stage of the study.

Results: A total of 295 answer sets were analyzed. Willingness to change behavior was higher in the serious game group (82% [119/145] versus 56% [84/150]; *P*<.001), with an odds ratio of 3.86 (95% CI 2.18-6.81; *P*<.001) after adjusting for professional category and baseline knowledge, using a mixed effects logistic regression model with LTCF as a random effect. For more than two-thirds (142/203) of the participants, the feeling of playing an important role against the epidemic was the most important factor explaining their willingness to change behavior. Most of the participants unwilling to change their behavior answered that they were already applying all the guidelines.

Conclusions: The serious game "Escape COVID-19" was more successful than standard IPC material in convincing LTCF employees to adopt COVID-19–safe IPC behavior.

International Registered Report Identifier (IRRID): RR2-10.2196/25595

(J Med Internet Res 2021;23(3):e27443) doi:10.2196/27443

KEYWORDS

COVID-19; transmission; serious game; infection prevention; health care worker; SARS-CoV-2; nursing homes; randomized controlled trial; long-term care facilities; impact; game; intention; control; elderly

Introduction

Background and Importance

Long-term care facilities (LTCFs) have been hit hard by the COVID-19 pandemic [1-4]. Although most LTCF residents are either old or frail and therefore more prone to complications and death following SARS-CoV-2 infection [5,6], other factors, such as shortages of human resources [7-9], must also be taken into account. This lack of resources can lead to suboptimal application of infection prevention and control (IPC) procedures [10], thus facilitating viral transmission. As even a single LTCF employee can contaminate a great number of colleagues and residents [11], and as rapid viral transmission between residents has been reported [12], it is essential to support LTCF employees [13] and ultimately help them adhere to IPC guidelines [14].

Even though most LTCF employees recognize the importance of preventing SARS-CoV-2 transmission, their motivation to act according to their awareness might be hampered by overwork [8,15], by too often updated and sometimes contradicting guidelines [16,17], and by mistrust in health care authorities [18]. Moreover, in spite of well-established evidence regarding specific IPC practices such as hand hygiene and use of personal protective equipment (PPE) [19], most health care workers (HCWs), including LTCF employees, only seldom apply them perfectly [20-22]. The current need for physical distance and the disruption of the regular continuing medical education courses has deteriorated this situation even further [23].

To enhance the communication of appropriate IPC guidelines and improve their application by HCWs and related staff who are regularly in contact with patients, we developed a web-based serious game called "Escape COVID-19" [24] using Nicholson's RECIPE for meaningful gamification [25]. Although electronic learning (e-learning) interventions might be ineffective in teaching complex technical procedures ([17,26]; Stuby et al, unpublished data, 2021), they are nevertheless useful in increasing knowledge and their use is generally associated with a high level of learner satisfaction [27-29]. As the probability of executing an action is strongly linked to the intention of performing it (according to the theory of planned behavior [30,31]), an engaging serious game could enhance the dissemination of essential IPC guidelines and encourage LTCF employees to change their behavior regarding IPC practices [29,32,33].

Objective

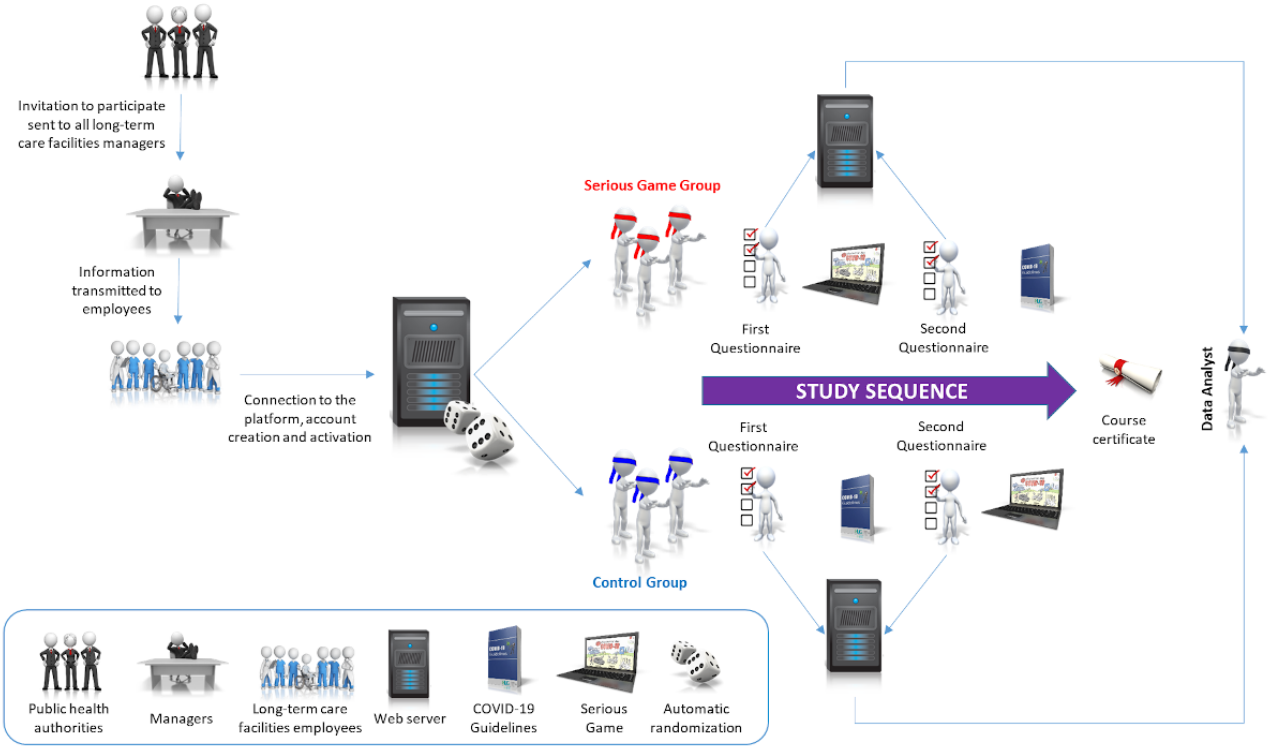
Our main objective was to assess the impact of "Escape COVID-19," a web-based serious game, on the intention of LTCF employees to change their IPC practices. Our secondary objective was to determine the reasons underlying the potential willingness to change IPC practices or lack thereof.

Methods

Study Design and Setting

This was a web-based, triple-blind (investigator, participants, data analyst), randomized controlled trial, which took place between November 5 and December 4, 2020. Its design has been published previously [34] and is summarized in Figure 1. A declaration of no objection was issued by our regional ethics committee (Req-2020-01262) as such projects do not fall within the scope of the Swiss federal law on human research [35]. This study was designed according to the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials) guidelines [36] and incorporates relevant elements from the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist [37].

Figure 1. Study design.

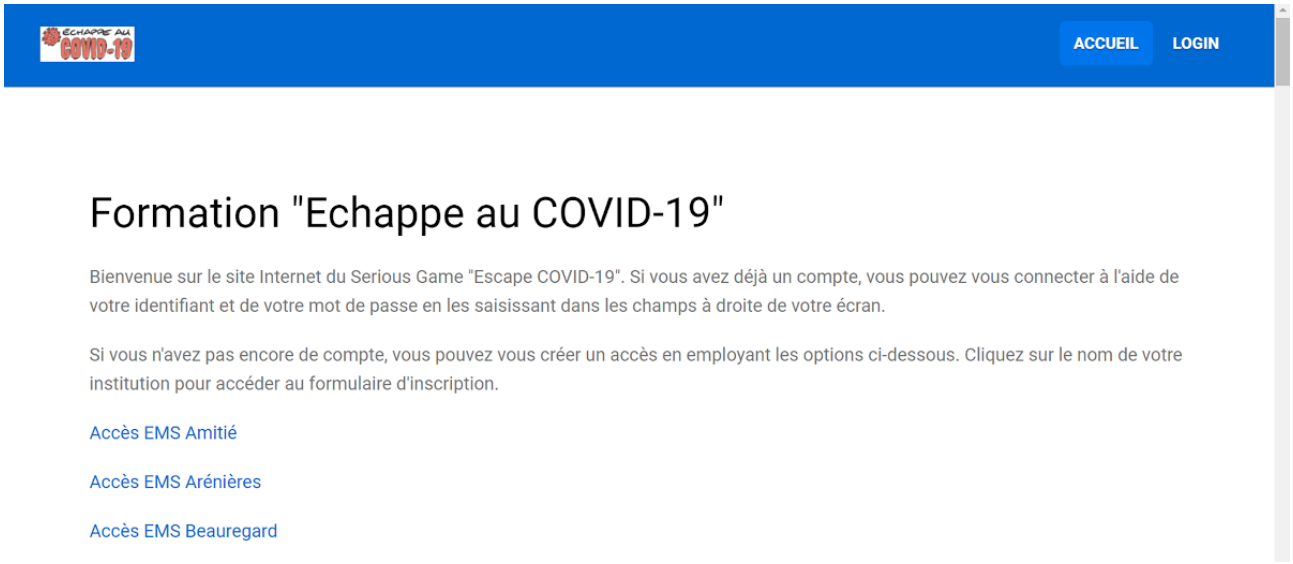


Having been informed by IPC specialists from the Geneva University Hospitals of the development of "Escape COVID-19," the public health authorities of Geneva, Switzerland, were interested in giving the employees of all LTCFs under their jurisdiction access to this serious game. They therefore provided us with a comprehensive list of all such facilities, but insisted on liaising with LTCF managers themselves. As LTCF employees other than HCWs can often be in contact with residents, we informed the representatives of the health care authorities that, for methodological and clinical relevance, all LTCF employees should be invited to participate, regardless of their professional status. We therefore provided them with an email template that described the objective of this study and the target population, and gave information regarding data protection (Multimedia Appendix 1). This template was sent by the public health authorities to all LTCF managers following a regular LTCF coordination meeting during which the characteristics of the study and the serious game were detailed. Participants, including LTCF managers, were unaware that they would be allocated to one of two different study arms but were required to provide consent (see below) and were given the approximate time required to complete the whole path (30 minutes). An email address that could be used to contact the investigators was also provided. The only incentive that participants were given was the promise of acquiring a course completion certificate at the end of the study path.

Online Platform

A specific web-based platform [38], hosted on a Swiss server, was created under the Joomla 3.9 content management system [39]. Only two authors (MS and L Suppan) had access to the administration console and data, which were stored on an encrypted MySQL-compatible database (MariaDB 5.5.5, MariaDB Foundation). Daily backups were performed throughout the study period. They were triggered by a cron job script and uploaded on a physically separate server through an encrypted connection. The front page displayed a comprehensive list of all LTCFs under Geneva health authorities' jurisdiction (Figure 2).

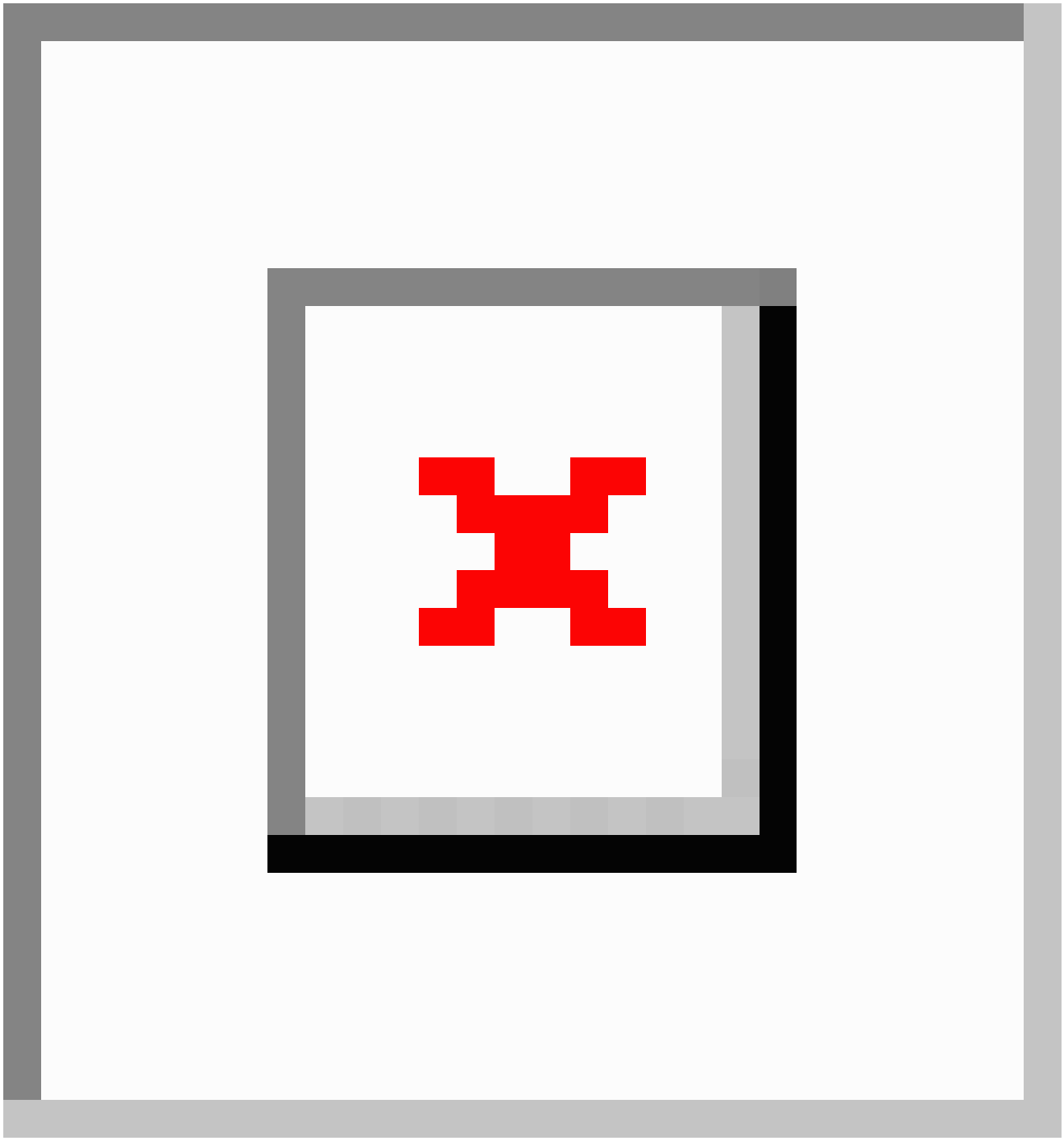
Figure 2. Front page of the study website (in French). Only the first three long-term care facilities are displayed.



Stratification was achieved by having LTCF employees click on a link specific to their institution. When they clicked on this link, participants were automatically randomized to one of the two study paths by GegaByte's Random Article module (GegaByte Technologies) [40]. This process was invisible to the end user and there was no way either participants or investigators could influence group allocation.

General information regarding the study, the need to register with a valid email address, and the need to allow the reception of emails coming from the website's domain was then displayed (Figure 3).

Figure 3. General information displayed (in French) before accessing the registration form.



After clicking on the "Access" button, a registration form, identical for both groups, was shown (Figure 4). This form was created using Membership Pro (Joomdonation) as this component enabled us to allocate users to specific groups, to create specific fields, and to disable the "Name" field [41]. Therefore, participants were only asked to enter their email address; they were not asked for any other personal information and were not required to give their first and last names.

Figure 4. Registration form (in French).

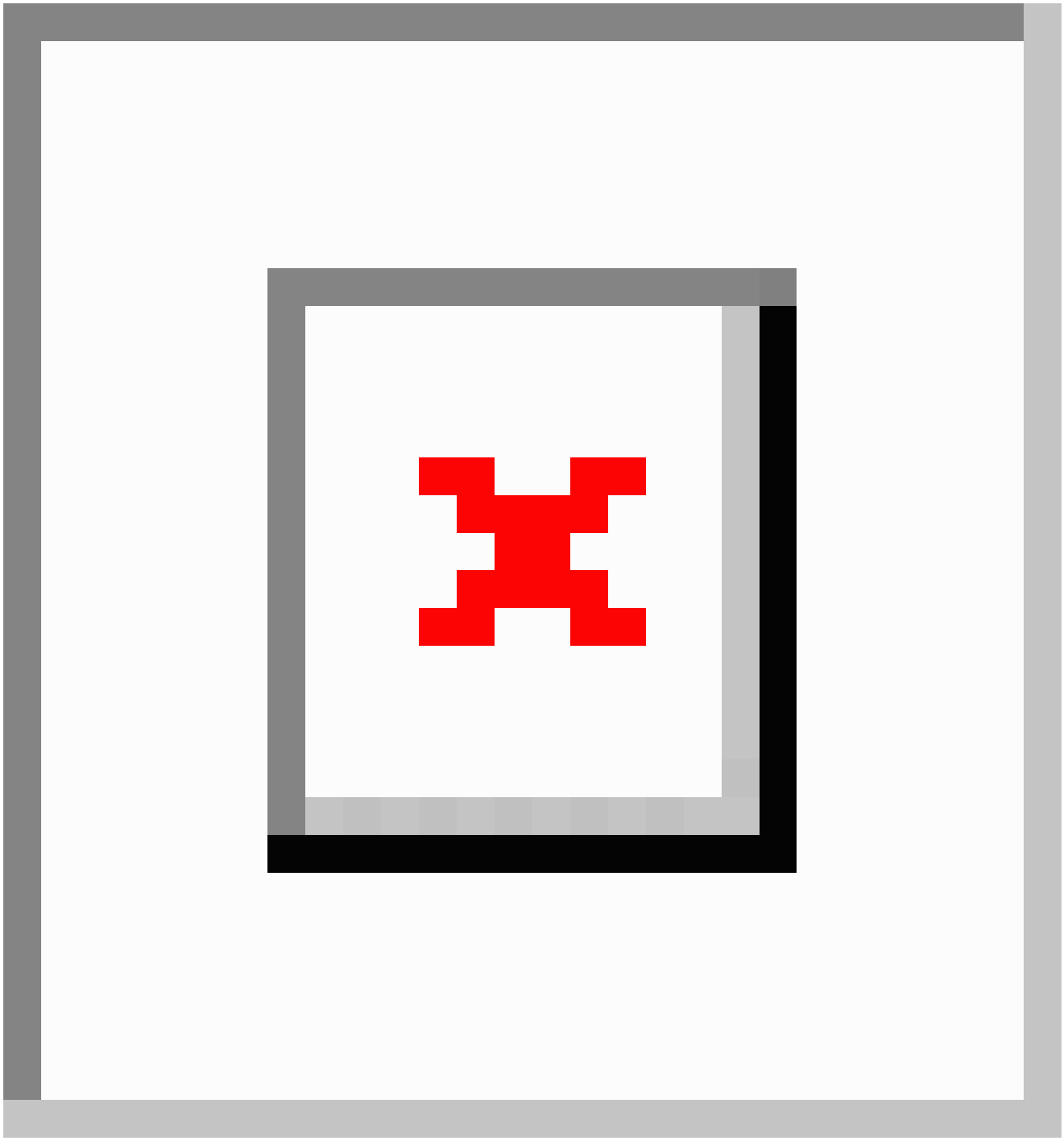
To avoid having LTCF employees registering under the wrong institution, LTCF-specific accreditations were created for each facility (Figure 4, "Accréditation" field). The full list of accreditations was sent to the public health authorities, who were informed they were to adapt our email template accordingly for each LTCF.

After filling in the registration form, participants were asked to activate their account using a specific link sent to the email address they had entered. This email also contained a reminder regarding data handling and security (Multimedia Appendix 2). Participants were informed that clicking on the activation link was considered as consent to participate in the study. After activating their account, participants were able to log in to the platform. Upon login, they were redirected to the first questionnaire (Multimedia Appendix 3) by the Redirect-on-Login 4.6 component (Pages-and-Items) [42].

This first questionnaire (Multimedia Appendix 3) was designed to assess the participants' baseline knowledge and to gather demographic data. It was developed using Community Surveys Pro (Corejoomla) [43], which enables a completeness check and allows for the use of branching logic. The number of initial questions was kept to a minimum and branching logic was used to try to limit attrition [44-47]. All multiple-choice and multiple-answer questions were mandatory and had to be answered before participants could move on to the next step. The answers could not be changed once a page had been completed.

After completing this first questionnaire, participants in the control group were shown a quick reminder of the most recent national guidelines published by the Federal Office of Public Health of the Swiss Confederation [48]. They were also given links to local IPC guidelines for health care professionals (Vigigerme) provided by the Geneva University Hospitals (Figure 5). After reviewing these guidelines, which are freely available on the internet [49], these participants were asked to complete the second questionnaire (Multimedia Appendix 4).

Figure 5. Control material (in French). The model was displayed for 10 seconds before providing links to infection prevention and control guidelines.

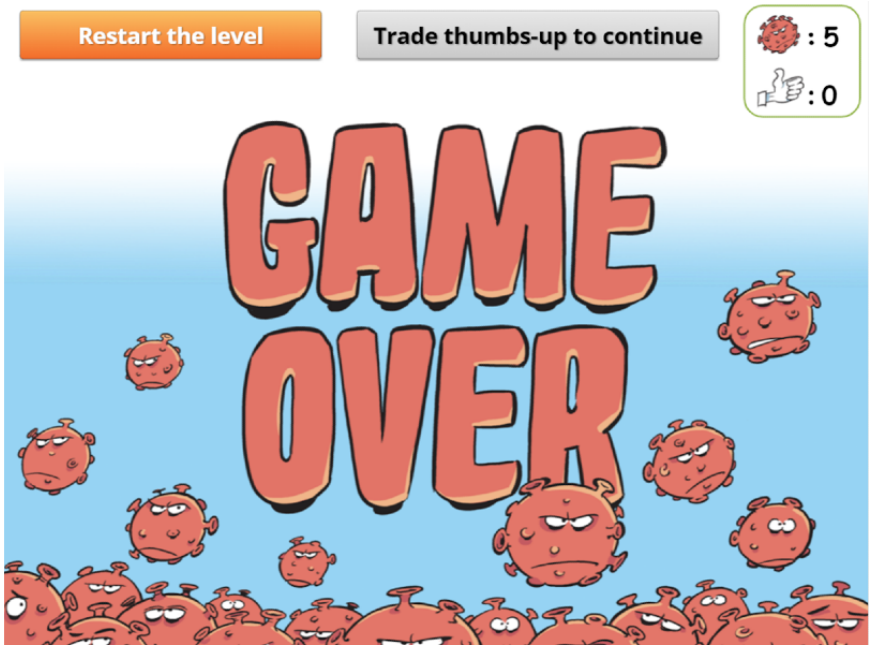


Conversely, “Escape COVID-19” (version 2.1.1) was launched once participants in the serious game group had completed the first questionnaire. The development of this serious game has previously been described [24]. Briefly, “Escape COVID-19” was first developed in French under Articulate Storyline 3 (Articulate Global) through multiple iterations by using the first three steps of the SERES framework [23,50,51] and Nicholson’s RECIPE for meaningful gamification [25]. Graphical elements were designed by Eric Buche [52] to lend a unique aspect to the game (Figures 6 and 7). The game has been fully translated in English and can be freely accessed on the internet [53]. The first and last author can be contacted to obtain Shareable Content Object Reference Model (SCORM) packages, which can be reused freely for research and educational purposes. Although the screen captures are presented in English to enhance their comprehension by an international readership, the original (French) versions can be seen in Multimedia Appendix 5.

Figure 6. Welcome screen of the Escape COVID-19 serious game.

This serious game includes four different levels, which create a meaningful narrative [25] by presenting recurrent and critical standard situations encountered by HCWs, including community and hospital exposures. Feedback is systematically given [54] to reinforce the expected behavior. Desirable behaviors and correct answers are rewarded by awarding a "thumbs-up," while viruses are accumulated whenever a wrong answer is given or an unwished-for behavior is performed. A "game over" screen is displayed (Figure 7) if the player gathers a total of 5 viruses. The player can then either restart the level or spend their thumbs-up to decrease the virus count (1:1).

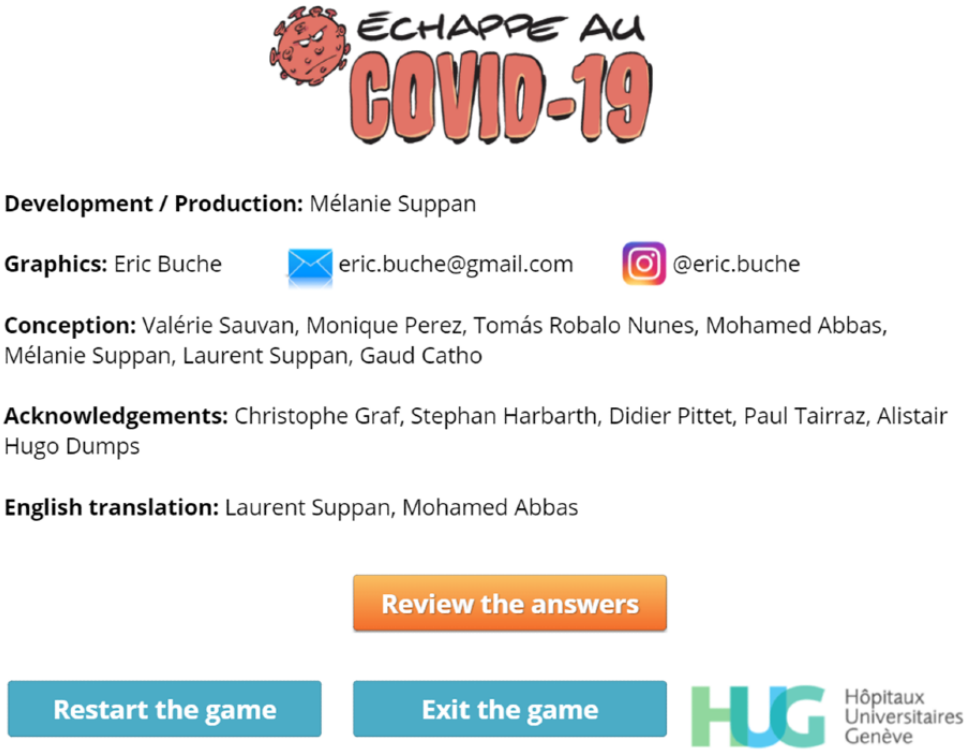
Figure 7. "Game over" screen of the Escape COVID-19 serious game.



After completing all four levels, participants were shown a link that enabled them to access the second questionnaire (Figure 8). This ensured that this questionnaire could only be answered by participants who had actually completed the game.

This second and last questionnaire was designed to assess whether the participants intended to change their IPC practices after completing their allocated set of learning materials. It was also developed using Community Surveys Pro, with branching logic used to try to limit attrition and completeness checks enabled to avoid missing data. After completing this last questionnaire, participants in the control group were granted access to the serious game, while those who belonged to the serious game group were shown the control materials. All participants could then access a certificate generation module that enabled them to obtain a course completion certificate. No time limit was set for either of the two questionnaires or for completing the serious game.

Figure 8. The only way participants could access the second questionnaire was by clicking on the "Exit the game" button.



During the whole study path, participants were allocated to specific user groups according to their progression. This was carried out by using Joomla's native Access Control List feature (PHP functions JUserHelper::addUserToGroup and JUserHelper::removeUserFromGroup, embedded using Sourcerer by Regular Labs [55]), and served two purposes. First, it allowed us to determine precisely at which step some participants elected to abandon the study. Second, it allowed participants who had been interrupted to be immediately redirected to the next incomplete step of the study when resuming.

Outcomes

The primary outcome was the proportion of LTCF employees who answered they were willing to change their IPC practices after seeing either the serious game or the control material. The secondary outcomes were the identification of factors associated with participant willingness to change their behavior, the reasons given by participants opposed to changing theirs [56], and the potential motivators which could have led them to change. Attrition was evaluated at each stage of the study.

We also assessed the IPC domains affected in participants who answered they were willing to change their IPC behavior, and whether these participants would modify their use of specific PPE items. In total, 13 questions were used to assess these latter outcomes. Therefore, conversely to the previous outcomes, answering questions related to IPC domains and PPE items were not compulsory to limit attrition.

Participants and Sample Size

Regardless of their professional status, all LTCF employees working in Geneva, Switzerland, who received the invitation and elected to participate were included in this study and represented a convenience sample. As the number of eligible employees was estimated to be approximately 4000 people, we hoped for a participation rate of around 20%. This would have allowed us to detect a difference of 10% at the .05 significance level with a power of 80% as we had calculated that 388 participants would have been needed in each group to detect such a difference.

Answer sets that had not been marked as completed by the computer system and those filled by participants other than LTCF employees (ie, government employees or employees of other institutions) were excluded.

Data Curation and Statistical Analysis

Data was exported by L Suppan in Microsoft Excel (Microsoft Corp; XLSX) and in comma-separated value (CSV) formats depending on the components before being imported, appended, and merged under Stata (StataCorp LLC). The groups were renamed using neutral names ("Atreides" and "Corrino"), the fields that could have led to the unblinding of the data analyst were removed, and the curated DTA file was transmitted to L Stuby, who used Stata (version 15.1) for statistical analysis. This investigator was not part of the serious game development team and did not coauthor the original publication, as we wanted to avoid any potential conflict of interest.

Univariable and multivariable logistic regression were used to assess the primary outcome, with adjustment performed according to prior knowledge (expressed as the percentage of correct answers), professional status, and facility. The log-linearity assumption was checked graphically and the goodness-of-fit was tested using the Hosmer-Lemeshow test. As randomization was stratified by center, we adjusted for this in the analysis by employing a random effects logistic regression model, using LTCF as a random effect. We tested the null hypothesis of absence of random effect using a chi-bar-square

test. We calculated the intraclass correlation coefficient (ICC) to quantify to what extent responses in a single LTCF were correlated.

Secondary outcomes were analyzed by assigning numerical values to the answers gathered through the use of Likert scales. As the domains potentially affected by a change in behavior were assessed using Likert scales ranging from 1 (not at all) to 6 (very much), the same values (ie, a score ranging from 1 to 6) were assigned to each item. The composite outcome was the sum of these 9 questions and was analyzed through univariable linear regression then adjusted by employing a mixed effects model, using LTCF as a random effect and the same adjustment variables as for the primary outcome.

The changes in the use of specific PPE items were assessed using 5-point Likert scales, ranging from "much less" to "much more." An odd number was decided upon to allow participants to give a neutral answer. Values ranging from –2 to +2 were therefore assigned to each answer, with positive values attributed to changes enhancing IPC behavior. A composite outcome was generated by summing up these values. We used the same statistical method as for the prior composite outcome. When computing composite outcomes, the same weight was applied to all questions. As a reduction in the use of N95 respirators can also be considered as desirable depending on the setting, a sensitivity analysis was performed by analyzing the composite outcome with and without this particular item. Each individual question based on a Likert scale was also analyzed separately and the results are presented graphically, either in the manuscript or in multimedia appendices. No imputation technique was used.

Descriptive statistics were used to describe the factors associated with willingness or refusal to change behavior. The chi-square test or—if the expected frequencies assumption was not met—the Fisher exact test were used to assess differences between groups.

The curated data file (with the groups renamed as "Control" and "Serious Game") is available on the Mendeley Data repository [57].

Results

During the course of the study, the public health authorities of Geneva requested that we create specific accesses for other institutions whose members were not part of the target population. A total of 652 accounts were created, out of which 569 (87.3%) were activated. After exclusion of accounts that had not completed the second questionnaire and of those belonging to institutions other than LTCFs, 295 answer sets were analyzed (Figure 9). Participant characteristics are detailed in Table 1.

Figure 9. Study flowchart.

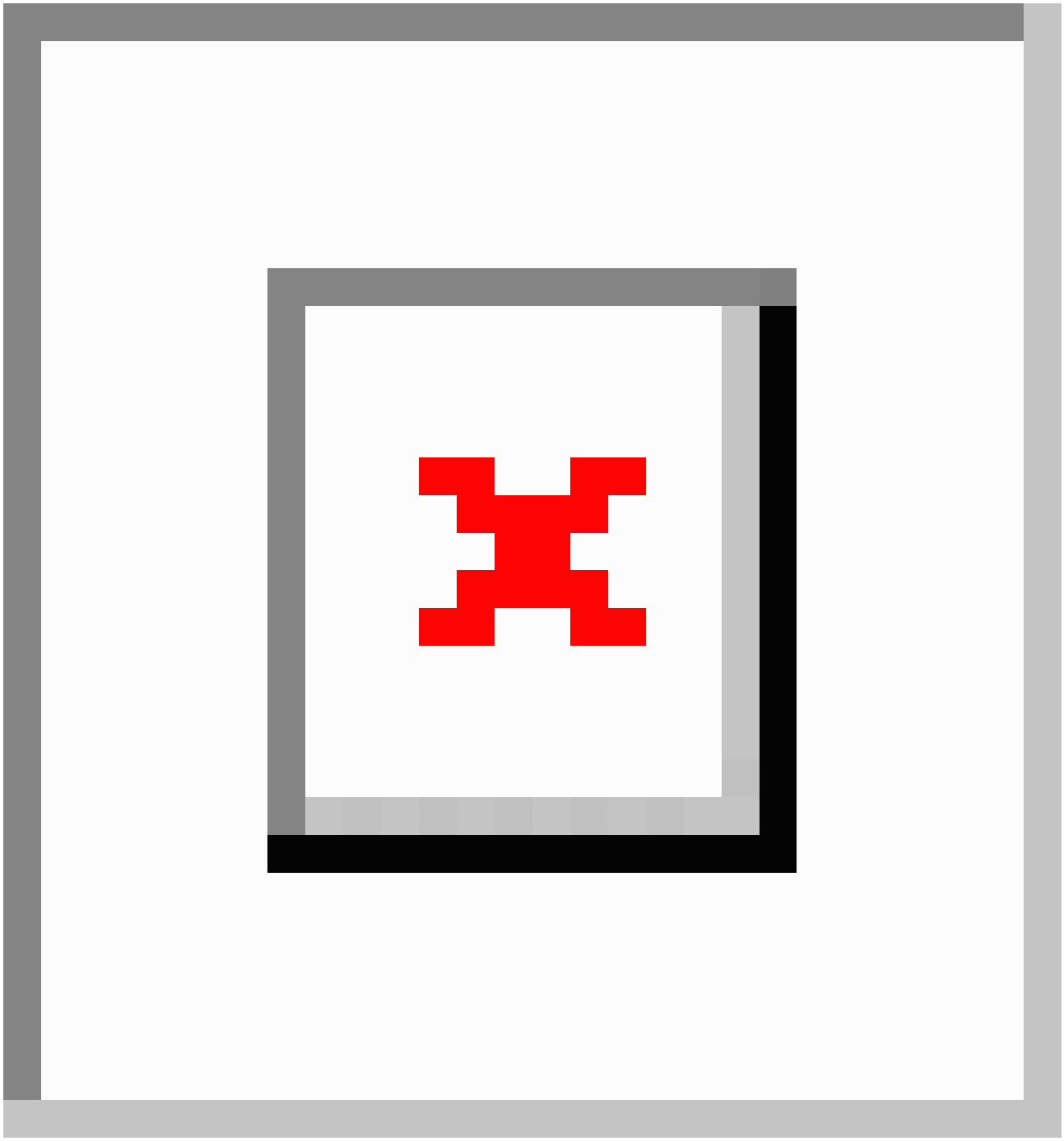


Table 1. Participant characteristics from 36 long-term care facilities in Geneva, Switzerland. Totals may not equal 100% due to rounding.

Participant characteristics	Control (n=150)	Serious game (n=145)
Gender, female, n (%)	118 (78.7)	111 (76.6)
Age (years), median (Q1-Q3)	45 (39-51)	43 (34-52)
Professional group, n (%)		
Nursing staff	83 (55.3)	83 (57.2)
Administrative/support staff	30 (20.0)	24 (16.6)
Physicians	3 (2.0)	3 (2.1)
Other	34 (22.7)	35 (24.1)
Detailed nursing staff status, n (%)		
Nurse	32 (38.6)	35 (41.2)
Health care assistant	36 (43.4)	34 (40.0)
Nurse assistant	5 (6.0)	11 (12.9)
Other	10 (12.1)	5 (5.9)
Non–health care professionals, n (%)		
Hospitality/catering	20 (58.8)	14 (40.0)
Animation	11 (32.4)	12 (34.3)
Other	3 (8.8)	9 (25.7)
Self-assessed frequency of patient contact for non–health care professionals, n (%)		
Very often	25 (39.1)	25 (41.0)
Quite often	19 (29.7)	12 (19.7)
Infrequently	15 (23.4)	22 (36.1)
Almost never	5 (7.8)	2 (3.3)
Years active in the health sector, median (Q1-Q3)	11 (6-22)	12 (4-21)
Baseline knowledge (percentage of correct answers), median (Q1-Q3)	40 (20-60)	40 (20-80)

Out of 36 LTCFs, only 8 provided 10 or more full answer sets, totaling 178 (60.3%) of all analyzed answers. Furthermore, one-third (n=12) of all LTCFs provided less than 5 full answer sets.

The willingness to change behavior was higher in the serious game group (82% [119/145], 95% CI 76%-88% versus 56% [84/150], 95% CI 48%-64%; *P*<.001), with an unadjusted odds ratio of 3.60 (95% CI 2.11-6.13; *P*<.001). After adjusting for professional category and baseline knowledge, using a random effects logistic regression model with LTCF as a random effect, the magnitude of the effect increased slightly, with an odds ratio of 3.86 (95% CI 2.18-6.81; *P*<.001). The effect was not significantly affected by professional category (*P*=.46) or baseline knowledge (*P*=.52). The ICC of 0.07 (95% CI 0.01-0.33) suggests little correlation of responses within individual LTCFs, although the chi-bar-square test showed that there was good evidence against the null hypothesis of no random effects (*P*=.046). A sensitivity analysis performed by excluding answers coming from LTCFs with <10 answers yielded an unadjusted odds ratio of 2.42 (95% CI 1.25-4.68; *P*=.009) and an adjusted odds ratio of 2.54 (95% CI 1.25-5.13; *P*=.01).

The factors underlying the willingness or lack thereof to change IPC behavior are detailed in Tables 2 and 3. The factors that could have led participants to change their behavior can be found in Table 4.

Table 2. Factors underlying the willingness to change infection prevention and control behavior.

Factors	Control (n=84), n (%)	Serious game (n=119), n (%)
The feeling of playing an important role in the common effort against the epidemic	60 (71.4)	82 (68.9)
The information given in the training material	55 (65.5)	74 (62.2)
The probability of infecting a relative	42 (50.0)	58 (48.7)
One should follow the procedures	38 (45.2)	57 (47.9)
Other	1 (1.2)	3 (2.5)

Table 3. Factors underlying the lack of willingness to change infection prevention and control behavior.

Factors	Control (n=66), n (%)	Serious game (n=26), n (%)
I already apply all these guidelines	62 (93.9)	23 (88.5)
This material was not in line with my situation	6 (9.1)	4 (15.4)
The material I have just seen was not helpful	0 (0.0)	0 (0.0)
I do not believe these measures to be useful	0 (0.0)	0 (0.0)
I disagree with these measures	0 (0.0)	0 (0.0)
Other	3 (4.6)	3 (11.5)

Table 4. Factors that could have brought about a change in infection prevention and control behavior.

Factors	Control (n=66), n (%)	Serious game (n=26), n (%)
A better understanding of the reasons underlying the recommendations	19 (28.8)	7 (26.9)
A greater probability of infecting a relative	12 (18.2)	8 (30.8)
The feeling of having an important role in the common effort against the epidemic	28 (42.4)	12 (46.2)
Another reason	19 (28.8)	5 (19.2)
Nothing—I could not have been convinced by any argument	2 (3.0)	2 (7.7)

Among the participants who answered that they were willing to change their IPC behavior after following the learning material, those in the control group exhibited a higher intensity in their willingness to adopt the recommended behaviors (4.88, 95% CI 0.56-9.22; *P*=.03). Adjustment for baseline knowledge, professional status, and LTCF slightly increased the effect, which did not change direction and remained statistically significant (4.98, 95% CI 0.85-9.10; *P*=.02). When analyzed separately, there were no significant differences between the control and intervention group in six IPC domains: not going to work if symptomatic (*P*=.07), protection from both colleagues and patients (*P*=.09), donning sequences with and without risk of aerosolization (*P*=.54 and *P*=.72, respectively), more frequently changing gloves (*P*=.26), and practicing hand hygiene (*P*=.33; Multimedia Appendix 6). However, participants in the control group felt significantly more concerned regarding workplace disinfection (*P*=.002; Figure 10), handling of the face mask (*P*=.02; Figure 11), and protecting themselves from asymptomatic people (*P*=.04) than participants in the intervention group (Figure 12).

Figure 10. Magnitude of the willingness to change workplace disinfection behavior.

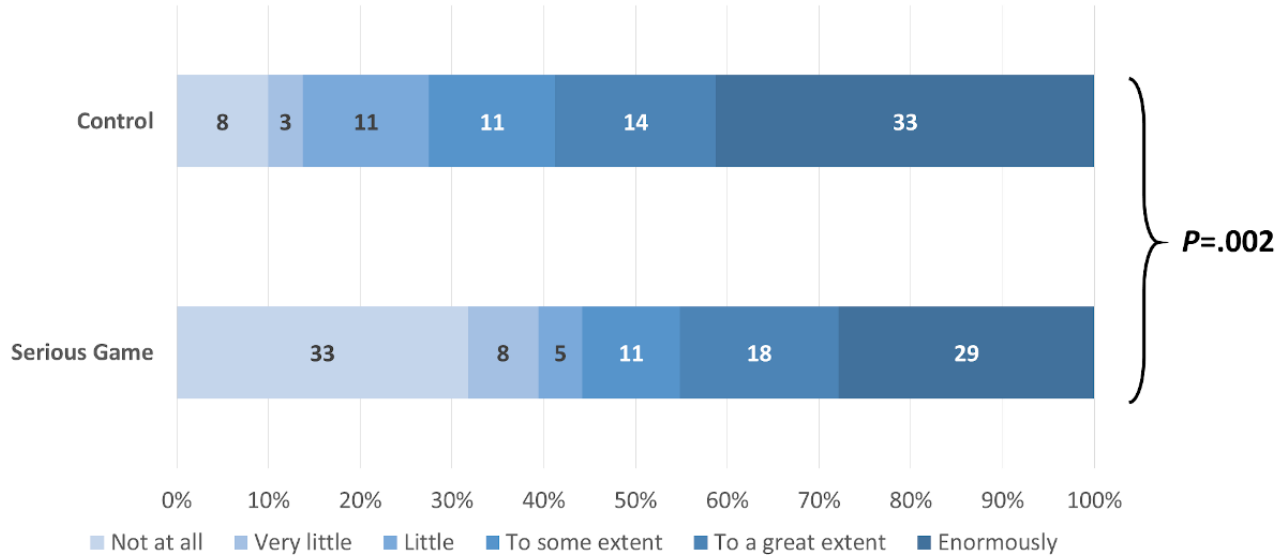


Figure 11. Magnitude of the willingness to change face mask handling behavior.

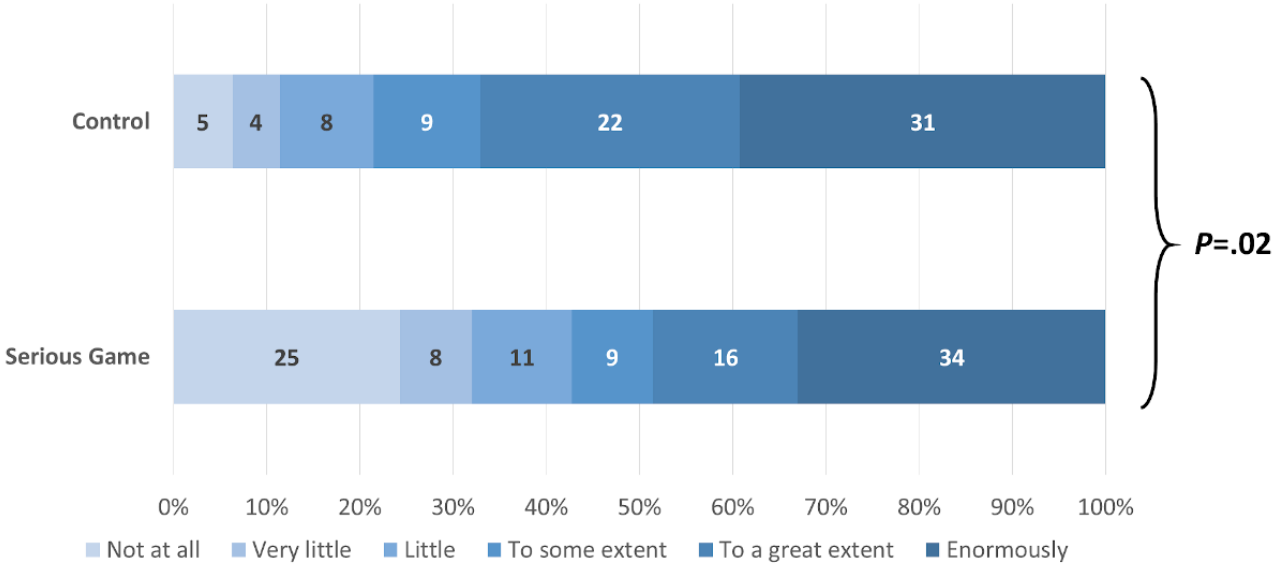
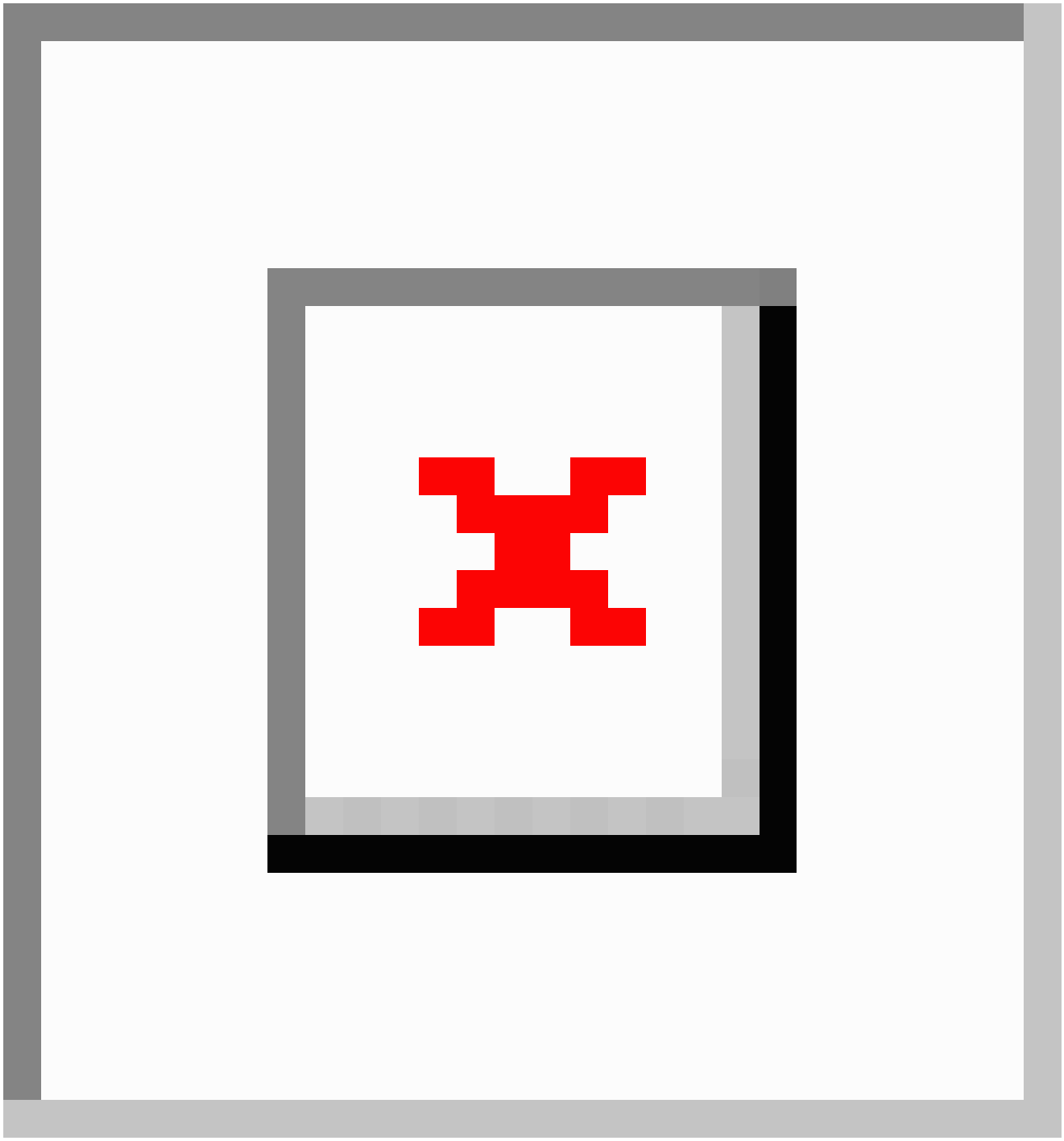


Figure 12. Magnitude of the willingness to protect oneself from asymptomatic people.



There was no overall difference between the groups in the intention of participants to change their use of specific PPE items after studying the IPC material (0.56, 95% CI –0.36 to 1.48; *P*=.23). Discarding the question about the use of N95 respirators did not affect this result (0.45, 95% CI –0.24 to 1.14; *P*=.20). There was no significant difference when PPE items were analyzed separately (surgical masks, *P*=.18; N95 respirators, *P*=.42; ocular protections, *P*=.22; gloves, *P*=.47; Multimedia Appendix 7).

Discussion

Principal Results

After following an online learning path, most participants reported that they were willing to change their IPC behavior. The serious game "Escape COVID-19" was, however, significantly more successful at inducing that change than the simple presentation of IPC guidelines. As this game was created through a theory-driven development process [24,25,50], its success at achieving the intended outcome reinforces the conclusions made by Gentry et al [29] in their recent systematic review, in which they called for further research using such methods.

Factors underlying the willingness to change IPC behavior were very similar between groups, with the feeling of playing an important role against the epidemic being most prominent, superseding even the information given in the training material. This could explain, at least in part, the success of the serious game over the standard IPC guidelines. Indeed, the "exposition" and "engagement" elements of Nicholson's RECIPE for meaningful gamification [25] were extensively used in the development of "Escape COVID-19" [24]. The exposition element is related to the creation of a meaningful narrative in the serious game. This was achieved by having the player go through steps they would usually encounter during the course of a regular work day, including informal times such as breaks or meals. The engagement element is two tiered, the first tier being linked to social engagement and the second to the concept of "flow." This latter concept relates to a progressive increase in the game's difficulty to avoid disinterest [58]. The former, social engagement, is usually achieved by the creation of multiplayer modes. Although "Escape COVID-19" lacks such a mode, it nevertheless takes this element into account by having the player make decisions that would, in real life, affect other people. To further strengthen the importance of these elements, analyzing the factors that could have motivated a change in IPC behavior revealed that the feeling of having an important role in the common effort against the epidemic was the leading factor that could have convinced participants to adapt their practices. This finding is in line with the driving force of the personal locus of control and the receptiveness to learning about and engaging in new behavior in the health care field [59].

The factors underlying the lack of willingness to change IPC behavior were also similar between groups. The vast majority of participants unwilling to change answered that they were already applying all the guidelines they had been presented with. As the serious game contained little material other than that presented in the standard IPC guidelines, it might therefore be surmised that the serious game was also more successful in conveying key IPC messages. Nevertheless, the magnitude of the intention to change IPC practices was significantly higher for some specific aspects in the fewer participants who were willing to change their IPC practices after following the control materials. Our hypothesis is that these participants might have been looking for information regarding some specific IPC aspects and were from the start highly motivated to change their practices according to relevant and up-to-date IPC guidelines, regardless of the way the information was presented. Another explanation is that, even though the serious game was more engaging, it was not designed to give in-depth explanations regarding specific IPC measures. This might have contributed to this difference, even though the proportion of participants answering that "a better understanding of the reasons underlying the recommendations" would have made them more willing to change their IPC practices was similar between groups.

No participant answered that they disagreed with the IPC guidelines or that they did not believe such measures to be useful. Given the current inclination for fake news [60-62] and conspiracy theories [63-65], this is rather reassuring, even more so as some participants were not HCWs but members of administrative, catering, and hospitality staff. More reassuring still is that the analysis of the factors that could have brought about a change in IPC behavior showed that only very few participants answered that they could not have been convinced by any argument.

Limitations and Strengths

This study has several limitations. First, even though the probability of executing an action is strongly linked to the intention of performing it, one can hardly be certain that LTCF employees claiming they are willing to change their IPC behavior will actually change it. Field observations would be necessary to ascertain this aspect, along with a different study design as both groups ultimately accessed the serious game in this study. Another important limitation is that we did not reach the sample size we had expected [34]. Indeed, while we had hoped for at least 800 participants, the actual number of accounts activated by LTCF employees was rather lower. This low figure raises many questions and hypotheses. Indeed, while at least a complete answer set was obtained for each individual LTCF, less than 10 complete answer sets were given by more than three-quarters (28/36) of all LTCFs under the jurisdiction of the public health authorities of Geneva, with almost half of those (12/28) giving less than 5 complete answer sets. It might therefore be assumed that, while all LTCF managers received the information regarding the study, many decided not to forward it to their employees. This was an unexpected finding; however, this study was not designed to assess the reasons underlying this decision. Hypotheses can however be drawn, some of which are more concerning than others. Among the least concerning, fear of overloading already overworked LTCF employees with information, an insufficient number of reminders, or the simple lack of regularly updated mailing lists could partly explain the low participation rate witnessed in many LTCFs. Moreover, some managers might have felt that the material, which originated from a university hospital, was not in line with their situation. However, this hypothesis is challenged by the fact that this impression, though asked for, was reported by 10 participants only. Lack of eHealth literacy is probably not to blame for the lack of participation. Indeed, LTCF employees increasingly use digital devices in the course of their work [66], and recent surveys have shown that eHealth literacy was rather high in HCWs [67]. Among the most disturbing hypotheses, a low level of concern of some LTCF managers, a potential mistrust of health authorities or the institution authoring the game, or even of IPC guidelines, cannot be ruled out. The will to avoid spreading information that could lead to an increase in the use of PPE items and, therefore, to an increase in material costs, seems unlikely. Regardless of the reason, the ability of health authorities to successfully convey critical messages by efficient vectors to LTCF employees should be assessed specifically to solve potential communication issues. The creation of IPC focal points in each LTCF is a path that could be explored.

Despite these limitations, this study also has some strengths, among which the fully automated randomization process, the triple-blinding, and the originality of the material could be mentioned. Finally, despite the lower-than-expected participation rate, the presence of a control group has enabled us to limit certain biases, as one could also hypothesize that only the most

motivated LTCF employees would have participated, thereby creating a selection bias and limiting the interpretation of our findings.

Perspectives

This serious game might not be equally effective in all populations, and IPC messages might differ from one region to another. By virtue of its flexible design, "Escape COVID-19" can be updated rather easily and should now be tested on other populations. It has been fully translated into English and is in the process of being translated into German and Italian to allow its deployment at the Swiss national level. To enhance its visibility, publicizing actions similar to those used to promote another recently developed serious game ("COVID-19 – Did You Know?) should be considered [68]. Currently, "Escape COVID-19" is freely available to play online [53], and the corresponding author can be contacted at any time to obtain a SCORM package in either French or English, while translation into other languages is pending.

Conclusion

Among LTCF employees, the serious game "Escape COVID-19" was more successful than standard IPC material in inspiring the willingness to adopt COVID-19–safe IPC behavior.

Acknowledgments

Many thanks to Mr Eric Buche, who designed the graphics used in the "Escape COVID-19" serious game. We also would like to extend our thanks to Monica Perez, Valérie Sauvan, and Tomás Robalo Nunes from the Geneva University Hospitals Infection Prevention and Control department, who participated in the development of "Escape COVID-19."

Conflicts of Interest

None declared.

Multimedia Appendix 1
Email template.
[PDF File (Adobe PDF File), 57 KB - jmir_v23i3e27443_app1.pdf]

Multimedia Appendix 2
Email received upon account creation.
[PNG File , 268 KB - jmir_v23i3e27443_app2.png]

Multimedia Appendix 3
First questionnaire.
[PDF File (Adobe PDF File), 132 KB - jmir_v23i3e27443_app3.pdf]

Multimedia Appendix 4
Second questionnaire.
[PDF File (Adobe PDF File), 105 KB - jmir_v23i3e27443_app4.pdf]

Multimedia Appendix 5
Original versions of the screen captures, in French.
[PDF File (Adobe PDF File), 553 KB - jmir_v23i3e27443_app5.pdf]

Multimedia Appendix 6
Graphical results of questions regarding infection prevention and control domains (based on the 6-point Likert scale).
[PDF File (Adobe PDF File), 533 KB - jmir_v23i3e27443_app6.pdf]

Multimedia Appendix 7
Graphical results of questions regarding specific personal protective equipment items (based on the 5-point Likert scale).
[PDF File (Adobe PDF File), 509 KB - jmir_v23i3e27443_app7.pdf]

Multimedia Appendix 8
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1214 KB - jmir_v23i3e27443_app8.pdf]

References

1. Szczerbińska K. Could we have done better with COVID-19 in nursing homes? Eur Geriatr Med 2020 Aug;11(4):639-643 [FREE Full text] [doi: 10.1007/s41999-020-00362-7] [Medline: 32676866]
2. Burki T. England and Wales see 20 000 excess deaths in care homes. Lancet 2020 May 23;395(10237):1602 [FREE Full text] [doi: 10.1016/S0140-6736(20)31199-5] [Medline: 32446403]
3. ECDC Public Health Emergency Team, Danis K, Fonteneau L, Georges S, Daniau C, Bernard-Stoecklin S, et al. High impact of COVID-19 in long-term care facilities, suggestion for monitoring in the EU/EEA, May 2020. Euro Surveill 2020 Jun;25(22):1 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.22.2000956] [Medline: 32524949]
4. Robalo Nunes T, Lebowitz D, Fraccaro M, Perez M, Vieux L, Abbas M, et al. Can long-term care facilities remain a coronavirus disease 2019 (COVID-19)-free bubble? An outbreak report. Infect Control Hosp Epidemiol 2021 Jan 11:1-2 [FREE Full text] [doi: 10.1017/ice.2020.1432] [Medline: 33427151]
5. O'Driscoll M, Ribeiro Dos Santos G, Wang L, Cummings DAT, Azman AS, Paireau J, et al. Age-specific mortality and immunity patterns of SARS-CoV-2. Nature 2021 Feb 02;590(7844):140-145. [doi: 10.1038/s41586-020-2918-0] [Medline: 33137809]
6. Mallapaty S. The coronavirus is most deadly if you are older and male - new data reveal the risks. Nature 2020 Sep 28;585(7823):16-17. [doi: 10.1038/d41586-020-02483-2] [Medline: 32860026]
7. Scales K. It's Time to Resolve the Direct Care Workforce Crisis in Long-Term Care. Gerontologist 2020 Aug 27:1 [FREE Full text] [doi: 10.1093/geront/gnaa116] [Medline: 32853357]
8. Van Houtven CH, DePasquale N, Coe NB. Essential Long-Term Care Workers Commonly Hold Second Jobs and Double- or Triple-Duty Caregiving Roles. J Am Geriatr Soc 2020 Aug 14;68(8):1657-1660 [FREE Full text] [doi: 10.1111/jgs.16509] [Medline: 32338767]
9. McGilton KS, Escrig-Pinol A, Gordon A, Chu CH, Zúñiga F, Sanchez MG, et al. Uncovering the Devaluation of Nursing Home Staff During COVID-19: Are We Fuelling the Next Health Care Crisis? J Am Med Dir Assoc 2020 Jul;21(7):962-965 [FREE Full text] [doi: 10.1016/j.jamda.2020.06.010] [Medline: 32674829]
10. McDonald MV, Brickner C, Russell D, Dowding D, Larson EL, Trifilio M, et al. Observation of Hand Hygiene Practices in Home Health Care. J Am Med Dir Assoc 2020 Sep 15:1 [FREE Full text] [doi: 10.1016/j.jamda.2020.07.031] [Medline: 32943340]
11. McMichael TM, Currie DW, Clark S, Pogosjans S, Kay M, Schwartz NG, Public Health–Seattle King County, Evergreen Health, CDC COVID-19 Investigation Team. Epidemiology of Covid-19 in a Long-Term Care Facility in King County, Washington. N Engl J Med 2020 May 21;382(21):2005-2011 [FREE Full text] [doi: 10.1056/NEJMoa2005412] [Medline: 32220208]
12. Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, Jacobs JR, Public Health–Seattle King County CDC COVID-19 Investigation Team. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. N Engl J Med 2020 May 28;382(22):2081-2090 [FREE Full text] [doi: 10.1056/NEJMoa2008457] [Medline: 32329971]
13. Werner RM, Hoffman AK, Coe NB. Long-Term Care Policy after Covid-19 - Solving the Nursing Home Crisis. N Engl J Med 2020 Sep 03;383(10):903-905. [doi: 10.1056/NEJMp2014811] [Medline: 32459918]
14. Abbas M, Robalo Nunes T, Martischang R, Zingg W, Iten A, Pittet D, et al. Nosocomial transmission and outbreaks of coronavirus disease 2019: the need to protect both patients and healthcare workers. Antimicrob Resist Infect Control 2021 Jan 06;10(1):7 [FREE Full text] [doi: 10.1186/s13756-020-00875-7] [Medline: 33407833]
15. White EM, Wetle TF, Reddy A, Baier RR. Front-line Nursing Home Staff Experiences During the COVID-19 Pandemic. J Am Med Dir Assoc 2021 Jan;22(1):199-203 [FREE Full text] [doi: 10.1016/j.jamda.2020.11.022] [Medline: 33321076]
16. Houghton C, Meskell P, Delaney H, Smalle M, Glenton C, Booth A, et al. Barriers and facilitators to healthcare workers' adherence with infection prevention and control (IPC) guidelines for respiratory infectious diseases: a rapid qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Apr 21;4:CD013582. [doi: 10.1002/14651858.CD013582] [Medline: 32315451]
17. Suppan L, Abbas M, Stuby L, Cottet P, Larribau R, Golay E, et al. Effect of an E-Learning Module on Personal Protective Equipment Proficiency Among Prehospital Personnel: Web-Based Randomized Controlled Trial. J Med Internet Res 2020 Aug 21;22(8):e21265 [FREE Full text] [doi: 10.2196/21265] [Medline: 32747329]
18. Cheung ATM, Parent B. Mistrust and inconsistency during COVID-19: considerations for resource allocation guidelines that prioritise healthcare workers. J Med Ethics 2021 Feb 26;47(2):73-77. [doi: 10.1136/medethics-2020-106801] [Medline: 33106381]
19. Rios P, Radhakrishnan A, Williams C, Ramkissoon N, Pham B, Cormack GV, et al. Preventing the transmission of COVID-19 and other coronaviruses in older adults aged 60 years and above living in long-term care: a rapid review. Syst Rev 2020 Sep 25;9(1):218 [FREE Full text] [doi: 10.1186/s13643-020-01486-4] [Medline: 32977848]
20. Fan J, Jiang Y, Hu K, Chen X, Xu Q, Qi Y, et al. Barriers to using personal protective equipment by healthcare staff during the COVID-19 outbreak in China. Medicine (Baltimore) 2020 Nov 25;99(48):e23310 [FREE Full text] [doi: 10.1097/MD.0000000000023310] [Medline: 33235088]
21. Michel-Kabamba N, Ngatu NR, Leon-Kabamba N, Katumbo-Mukemo A, Mukuku O, Ngoyi-Mukonkole J, et al. Occupational COVID-19 Prevention among Congolese Healthcare Workers: Knowledge, Practices, PPE Compliance, and Safety

Imperatives. Trop Med Infect Dis 2020 Dec 30;6(1):6 [FREE Full text] [doi: 10.3390/tropicalmed6010006] [Medline: 33396599]
22. Maltezou H, Dedoukou X, Tsonou P, Tseroni M, Raftopoulos V, Pavli A, et al. Hospital factors associated with SARS-CoV-2 infection among healthcare personnel in Greece. J Hosp Infect 2021 Mar;109:40-43 [FREE Full text] [doi: 10.1016/j.jhin.2020.10.010] [Medline: 33169676]
23. Suppan M, Gartner B, Golay E, Stuby L, White M, Cottet P, et al. Teaching Adequate Prehospital Use of Personal Protective Equipment During the COVID-19 Pandemic: Development of a Gamified e-Learning Module. JMIR Serious Games 2020 Jun 12;8(2):e20173 [FREE Full text] [doi: 10.2196/20173] [Medline: 32516115]
24. Suppan M, Catho G, Robalo Nunes T, Sauvan V, Perez M, Graf C, et al. A Serious Game Designed to Promote Safe Behaviors Among Health Care Workers During the COVID-19 Pandemic: Development of "Escape COVID-19". JMIR Serious Games 2020 Dec 03;8(4):e24986 [FREE Full text] [doi: 10.2196/24986] [Medline: 33242312]
25. Nicholson S. A RECIPE for Meaningful Gamification. In: Reiners T, Wood L, editors. Gamification in Education and Business. Cham: Springer; 2015:1-20.
26. Suppan L, Stuby L, Gartner B, Larribau R, Iten A, Abbas M, et al. Impact of an e-learning module on personal protective equipment knowledge in student paramedics: a randomized controlled trial. Antimicrob Resist Infect Control 2020 Nov 10;9(1):185 [FREE Full text] [doi: 10.1186/s13756-020-00849-9] [Medline: 33168097]
27. Koka A, Suppan L, Cottet P, Carrera E, Stuby L, Suppan M. Teaching the National Institutes of Health Stroke Scale to Paramedics (E-Learning vs Video): Randomized Controlled Trial. J Med Internet Res 2020 Jun 09;22(6):e18358 [FREE Full text] [doi: 10.2196/18358] [Medline: 32299792]
28. Suppan M, Stuby L, Carrera E, Cottet P, Koka A, Assal F, et al. Asynchronous Distance Learning of the National Institutes of Health Stroke Scale During the COVID-19 Pandemic (E-Learning vs Video): Randomized Controlled Trial. J Med Internet Res 2021 Jan 15;23(1):e23594 [FREE Full text] [doi: 10.2196/23594] [Medline: 33428581]
29. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, et al. Serious Gaming and Gamification Education in Health Professions: Systematic Review. J Med Internet Res 2019 Mar 28;21(3):e12994 [FREE Full text] [doi: 10.2196/12994] [Medline: 30920375]
30. Regard S, Rosa D, Suppan M, Giangaspero C, Larribau R, Niquille M, et al. Evolution of Bystander Intention to Perform Resuscitation Since Last Training: Web-Based Survey. JMIR Form Res 2020 Nov 30;4(11):e24798 [FREE Full text] [doi: 10.2196/24798] [Medline: 33252342]
31. Ajzen I. From Intentions to Actions: A Theory of Planned Behavior. In: Kuhl J, Beckmann J, editors. Action Control. Berlin, Heidelberg: Springer; 1985:11-39.
32. Drummond D, Hadchouel A, Tesnière A. Serious games for health: three steps forwards. Adv Simul (Lond) 2017 Feb 4;2(1):3 [FREE Full text] [doi: 10.1186/s41077-017-0036-3] [Medline: 29450004]
33. Kortteisto T, Kaila M, Komulainen J, Mäntyranta T, Rissanen P. Healthcare professionals' intentions to use clinical guidelines: a survey using the theory of planned behaviour. Implement Sci 2010 Jun 29;5(1):51 [FREE Full text] [doi: 10.1186/1748-5908-5-51] [Medline: 20587021]
34. Suppan L, Abbas M, Catho G, Stuby L, Regard S, Harbarth S, et al. Impact of a Serious Game on the Intention to Change Infection Prevention and Control Practices in Nursing Homes During the COVID-19 Pandemic: Protocol for a Web-Based Randomized Controlled Trial. JMIR Res Protoc 2020 Dec 15;9(12):e25595 [FREE Full text] [doi: 10.2196/25595] [Medline: 33296329]
35. Swiss Confederation. SR 810.30 Federal Act of 30 September 2011 on Research involving Human Beings (Human Research Act, HRA). 2011. URL: https://www.admin.ch/opc/en/classified-compilation/20061313/index.html [accessed 2020-05-12]
36. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
37. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
38. Escape COVID-19 Online Platform. URL: https://covid-escape.anesth.ch [accessed 2020-11-04]
39. Joomla Content Management System. URL: https://www.joomla.org/ [accessed 2020-09-30]
40. Gegabyte Random Article. Gegabyte Technologies. URL: https://www.gegabyte.org/downloads/joomla-extensions/joomla3/modules/291-random-article [accessed 2020-11-03]
41. Membership Pro. Joomdonation. URL: https://joomdonation.com/joomla-extensions/membership-pro-joomla-membership-subscription.html [accessed 2020-11-03]
42. Redirect-on-Login. Pages-and-Items. URL: https://www.pages-and-items.com/extensions/redirect-on-login [accessed 2020-11-03]
43. Corejoomla. Community Surveys Pro. URL: https://www.corejoomla.com/products/community-surveys.html [accessed 2020-11-03]
44. Hochheimer CJ, Sabo RT, Perera RA, Mukhopadhyay N, Krist AH. Identifying Attrition Phases in Survey Data: Applicability and Assessment Study. J Med Internet Res 2019 Aug 23;21(8):e12811 [FREE Full text] [doi: 10.2196/12811] [Medline: 31444875]

45. McCambridge J, Kalaitzaki E, White IR, Khadjesari Z, Murray E, Linke S, et al. Impact of length or relevance of questionnaires on attrition in online trials: randomized controlled trial. J Med Internet Res 2011 Nov 18;13(4):e96 [FREE Full text] [doi: 10.2196/jmir.1733] [Medline: 22100793]
46. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
47. Hochheimer CJ, Sabo RT, Krist AH, Day T, Cyrus J, Woolf SH. Methods for Evaluating Respondent Attrition in Web-Based Surveys. J Med Internet Res 2016 Nov 22;18(11):e301 [FREE Full text] [doi: 10.2196/jmir.6342] [Medline: 27876687]
48. Coronavirus - Protect yourself and others - Current situation. Federal Office of Public Health. URL: https://foph-coronavirus.ch/ [accessed 2020-11-03]
49. SARS-CoV-2 - Fiche Vigigerme. Hôpitaux Universitaires de Genève. URL: https://vigigerme.hug.ch/fiche-vigigerme/sars-cov-2 [accessed 2020-11-03]
50. Verschueren S, Buffel C, Vander Stichele G. Developing Theory-Driven, Evidence-Based Serious Games for Health: Framework Based on Research Community Insights. JMIR Serious Games 2019 May 02;7(2):e11565 [FREE Full text] [doi: 10.2196/11565] [Medline: 31045496]
51. Verschueren S, van Aalst J, Bangels A, Toelen J, Allegaert K, Buffel C, et al. Development of CliniPup, a Serious Game Aimed at Reducing Perioperative Anxiety and Pain in Children: Mixed Methods Study. JMIR Serious Games 2019 Jun 01;7(2):e12429 [FREE Full text] [doi: 10.2196/12429] [Medline: 31199333]
52. Eric Buche - Instagram. URL: https://www.instagram.com/eric.buche/ [accessed 2020-11-03]
53. Escape COVID-19 Serious Game. URL: https://escape-covid19.ch [accessed 2021-01-20]
54. Kopp V, Stark R, Fischer M. Fostering diagnostic knowledge through computer-supported, case-based worked examples: effects of erroneous examples and feedback. Med Educ 2008 Aug;42(8):823-829. [doi: 10.1111/j.1365-2923.2008.03122.x] [Medline: 18564096]
55. Sourcerer. Regular Labs. URL: https://regularlabs.com/extensions/sourcerer [accessed 2020-11-03]
56. Tummers L, Steijn B, Bekkers V. Explaining the willingness of public professionals to implement public policies: content, context, and personality characteristics. Public Adm 2012;90(3):716-736. [doi: 10.1111/j.1467-9299.2011.02016.x]
57. Suppan L, Stuby L, Suppan M. Impact of Escape COVID-19 on long-term care facilities employees dataset. Mendeley Data. 2021. URL: https://data.mendeley.com/datasets/7cjmmntf6j/1 [accessed 2021-03-11]
58. Csikszentmihalyi M. Finding Flow: The Psychology Of Engagement With Everyday Life. New York: Basic Books; 1997.
59. Wallston K. Control Beliefs: Health Perspectives. In: Wright JD, editor. International Encyclopedia of the Social & Behavioral Sciences (Second Edition). Amsterdam: Elsevier; 2015:819-821.
60. O'Connor C, Murphy M. Going viral: doctors must tackle fake news in the covid-19 pandemic. BMJ 2020 Apr 24;369:m1587. [doi: 10.1136/bmj.m1587] [Medline: 32332066]
61. Ahmad AR, Murad HR. The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study. J Med Internet Res 2020 May 19;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]
62. Rovetta A, Bhagavathula AS. Global Infodemiology of COVID-19: Analysis of Google Web Searches and Instagram Hashtags. J Med Internet Res 2020 Aug 25;22(8):e20673 [FREE Full text] [doi: 10.2196/20673] [Medline: 32748790]
63. Ahmed W, López Seguí F, Vidal-Alaball J, Katz MS. COVID-19 and the. J Med Internet Res 2020 Oct 05;22(10):e22374 [FREE Full text] [doi: 10.2196/22374] [Medline: 32936771]
64. Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. J Med Internet Res 2020 May 06;22(5):e19458 [FREE Full text] [doi: 10.2196/19458] [Medline: 32352383]
65. Pickles K, Cvejic E, Nickel B, Copp T, Bonner C, Leask J, et al. COVID-19 Misinformation Trends in Australia: Prospective Longitudinal National Survey. J Med Internet Res 2021 Jan 07;23(1):e23805 [FREE Full text] [doi: 10.2196/23805] [Medline: 33302250]
66. Wilson R, Small J. Care Staff Perspectives on Using Mobile Technology to Support Communication in Long-Term Care: Mixed Methods Study. JMIR Nursing 2020 Sep 29;3(1):e21881. [doi: 10.2196/21881]
67. Do BN, Tran TV, Phan DT, Nguyen HC, Nguyen TTP, Nguyen HC, et al. Health Literacy, eHealth Literacy, Adherence to Infection Prevention and Control Procedures, Lifestyle Changes, and Suspected COVID-19 Symptoms Among Health Care Workers During Lockdown: Online Survey. J Med Internet Res 2020 Nov 12;22(11):e22894 [FREE Full text] [doi: 10.2196/22894] [Medline: 33122164]
68. Gaspar J, Lage E, Silva FD, Mineiro E, Oliveira ID, Oliveira I, et al. A Mobile Serious Game About the Pandemic (COVID-19 - Did You Know?): Design and Evaluation Study. JMIR Serious Games 2020 Dec 22;8(4):e25226 [FREE Full text] [doi: 10.2196/25226] [Medline: 33301416]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
CONSORT: Consolidated Standards of Reporting Trials
HCW: health care worker

ICC: intraclass correlation coefficient
IPC: infection prevention and control
LTCF: long-term care facility
PPE: personal protective equipment
SCORM: Shareable Content Object Reference Model

Edited by C Basch; submitted 01.02.21; peer-reviewed by L Abensur Vuillaume, E Tashkandi; comments to author 06.03.21; revised version received 07.03.21; accepted 07.03.21; published 25.03.21.

Please cite as:
Suppan M, Abbas M, Catho G, Stuby L, Regard S, Achab S, Harbarth S, Suppan L
Impact of a Serious Game (Escape COVID-19) on the Intention to Change COVID-19 Control Practices Among Employees of Long-term Care Facilities: Web-Based Randomized Controlled Trial
J Med Internet Res 2021;23(3):e27443
URL: https://www.jmir.org/2021/3/e27443
doi:10.2196/27443
PMID:33685854

Original Paper

Relationship Between Coronavirus-Related eHealth Literacy and COVID-19 Knowledge, Attitudes, and Practices among US Adults: Web-Based Survey Study

Lawrence An[1,2], MD; Elizabeth Bacon[1], MPH; Sarah Hawley[1,2,3], PhD; Penny Yang[1], BSc; Daniel Russell[4], PhD; Scott Huffman[4], PhD; Ken Resnicow[1,5], PhD

[1]Center for Health Communications Research, Rogel Cancer Center, University of Michigan, Ann Arbor, MI, United States

[2]Division of General Medicine, School of Medicine University of Michigan, Ann Arbor, MI, United States

[3]VA Center for Clinical Management Research, VA Ann Arbor Healthcare System, Ann Arbor, MI, United States

[4]Google, Mountain View, CA, United States

[5]Department of Health Behavior & Health Education, University of Michigan School of Public Health, Ann Arbor, MI, United States

Corresponding Author:
Lawrence An, MD
Center for Health Communications Research
Rogel Cancer Center
University of Michigan
North Campus Research Complex, Building 16
2800 Plymouth Rd
Ann Arbor, MI, 48109
United States
Phone: 1 734 763 6099
Email: lcan@med.umich.edu

Abstract

Background: During a global pandemic, it is critical that the public is able to rapidly acquire new and accurate health information. The internet is a major source of health information. eHealth literacy is the ability of individuals to find, assess, and use health information available on the internet.

Objective: The goals of this study were to assess coronavirus-related eHealth literacy and examine the relationship between eHealth literacy and COVID-19−related knowledge, attitudes, and practices (KAPs).

Methods: We conducted a web-based survey of a representative sample of 1074 US adults. We adapted the 8-item eHealth Literacy Scale to develop the Coronavirus-Related eHealth Literacy Scale (CoV-eHEALS) to measure COVID-19−related knowledge, conspiracy beliefs, and adherence to protective behaviors (eg, wearing facial masks and social distancing). Our analyses identified sociodemographic associations with the participants' CoV-eHEALS scores and an association between the CoV-eHEALS measure and COVID-19 KAPs.

Results: The internal consistency of the adapted CoV-eHEALS measure was high (Cronbach α=.92). The mean score for the CoV-eHEALS was 29.0 (SD 6.1). A total of 29% (306/1074) of the survey participants were classified as having low coronavirus-related eHealth literacy (CoV-eHEALS score <26). Independent associations were found between CoV-eHEALS scores and ethnicity (standardized β=–.083, *P*=.016 for Black participants) and education level (standardized β=–.151, *P*=.001 for participants with high-school education or lower). Controlling for demographic characteristics, CoV-eHEALS scores demonstrated positive independent associations with knowledge (standardized β=.168, *P*<.001) and adherence to protective behaviors (standardized β=.241, *P*<.001) and a negative association with conspiracy beliefs (standardized β=–.082, *P*=.009).

Conclusions: This study provides an estimate of coronavirus-related eHealth literacy among US adults. Our findings suggest that a substantial proportion of US adults have low coronavirus-related eHealth literacy and are thus at a greater risk of lower and less-protective COVID-19 KAPs. These findings highlight the need to assess and address eHealth literacy as part of COVID-19 control efforts. Potential strategies include improving the quality of health information about COVID-19 available on the internet, assisting or simplifying web-based search for information about COVID-19, and training to improve general or coronavirus-specific search skills.

(J Med Internet Res 2021;23(3):e25042) doi:10.2196/25042

KEYWORDS

internet; digital health; eHealth; eHealth literacy; coronavirus; COVID-19; knowledge; conspiracy beliefs; protective behaviors; social distancing; survey; health communication; attitude; behavior

Introduction

During a global pandemic, it is critical that members of the public are able to rapidly acquire new and accurate health information [1-4]. This includes information about the causative agent, transmission and course of the disease, as well as prevention and treatment [5,6]. It is also important to be able to avoid misinformation that might discourage protective behaviors or even encourage actions that lead to self-harm. This includes specific mistaken beliefs (eg, the new disease is no more serious than common existing infections) and also broader conspiracy theories (ie, unsubstantiated and implausible assertions that hidden forces in control of the society have created or are using the pandemic to extend their authority) that can interfere with or undermine individual or public responses to the pandemic. Research continues to identify deficits in the public's knowledge of key facts regarding the current COVID-19 pandemic, particularly among underserved groups that bear a disproportionate burden in terms of the number of COVID-19 cases and deaths [7-10]. The prevalence and negative impact of misinformation and the spread of conspiracy beliefs during the current pandemic (eg, spread of the "Plandemic" viral video) have also been increasingly recognized [6,11-14].

The internet has become a major source of health information for the public [15-19]. eHealth literacy refers to the ability of individuals to find, assess, and effectively use health information available on the internet [20]. The eHealth Literacy Scale (eHEALS) is one of the most commonly used measures to assess eHealth literacy and has been shown to be reliable and valid across a range of health conditions and populations [21-32]. eHealth literacy may be particularly important during the current pandemic because some widespread disease-control strategies (eg, social distancing) can limit in-person contact and reduce transmission of key information through these channels. Several authors have highlighted the importance of considering eHealth literacy in the response to the current global COVID-19 pandemic [1-4]. A prior study has demonstrated an association between higher receipt of information about COVID-19 available on the internet and increased engagement in personal protective behavior against COVID-19, such as handwashing, wearing facial masks, and avoiding social gatherings [33].

The goals of this study were to: (1) understand the ability of individuals to identify, assess, and effectively utilize health information about the coronavirus available on the internet and (2) determine how this ability might be associated with COVID-19–related knowledge, attitudes, and practices (KAPs). We performed a web-based survey of a nationally representative sample of US adults to achieve these goals.

Methods

Measures

For each of the following measures, specific survey items were presented to the survey respondents in random order.

Coronavirus-Related eHealth Literacy

We slightly modified items from the well-established eHEALS [20] to focus specifically on health information available on the internet *about the coronavirus*. The resulting 8-item measure assesses an individual's self-rated ability (answered on a 5-point Likert scale) to use the internet to find and utilize health information about the coronavirus. The specific items of this coronavirus-related eHEALS (CoV-eHEALS) were as follows:

1. I know what health resources about coronavirus are available on the internet.
2. I know where to find helpful health resources about coronavirus on the internet.
3. I know how to find helpful health resources about coronavirus on the internet.
4. I know how to use the internet to answer my questions about my health and coronavirus.
5. I know how to use the health information about coronavirus I find on the internet to help me.
6. I have the skills I need to evaluate the health resources about coronavirus I find on the internet.
7. I can tell high-quality health resources from low-quality health resources about coronavirus on the internet.
8. I feel confident in using information about coronavirus from the internet to make health decisions.

An overall CoV-eHEALS score was computed based on the sum of the scores for each item (range 8-40). The internal consistency of the CoV-eHEALS measure was 0.92, and this was not improved by the deletion of any specific item. Some prior studies have also defined cut-off points to characterize respondents as having low versus high eHealth literacy [22,27,29]. Consistent with this work, we categorized respondents, based on their total CoV-eHEALS score, as having low (score <26) or high (score ≥26) coronavirus-related eHealth literacy.

COVID-19 KAPs

We assessed the survey respondents' COVID-19 KAPs by using the following measures.

Knowledge

We created a 7-item scale based on common key facts related to COVID-19, recognized as of May 2019 [5,6]. Each item was answered on a 5-point scale ranging from "Definitely false" to "Definitely true." The specific items of this scale were as follows:

1. Coronavirus can be easily spread from one person to another.
2. Many thousands of people have died from coronavirus.
3. A vaccine is not yet available for the coronavirus.
4. Most people already have immunity to coronavirus.
5. Symptoms of coronavirus are always visible.
6. There are effective treatments for coronavirus that can cure most people.
7. Having coronavirus is about as dangerous as having the flu.

After reverse-coding of items 4-7, we created an overall knowledge score based on a mean of the scores for each item (range 1-5). The internal consistency of this knowledge measure was 0.78, and this was not improved by the deletion of any specific item.

Conspiracy Beliefs

We developed a brief, 3-item scale based on prior studies on COVID-19 and other health issues [14,34]. The scale is intended to measure conspiracy beliefs regarding the coronavirus rather than a generalized conspiracy trait or worldview. Each item was answered along a 5-point continuum ranging from "Definitely false" to "Definitely true." The specific items of this scale were as follows:

1. The real truth about coronavirus is being kept from the public.
2. People in power are using coronavirus as an excuse to monitor and control the public.
3. The media is making coronavirus seem more dangerous that it really is.

We computed a mean of the response to these 3 items to create a conspiracy score (range 1-5). The internal consistency of this conspiracy measure was 0.74, and this was not improved by the deletion of any specific item.

Protective Behavior Adherence Score

We examined the frequency of 7 self-reported behaviors practiced by the survey respondents over the past week, all of which are recommended for reducing the risk of transmitting and/or acquiring COVID-19 [35]. Each item was answered on a 5-point continuum: "Rarely or never," "Some of the time," "Most of the time," "Almost all of the time," and "All of the time." Our measure shares many topics in common to a recently described measure of COVID-19 infection prevention behaviors [36]. The specific items of this scale were as follows:

1. Avoiding touching my face.
2. Keeping my hands clean (eg, washing longer with soap and water, using hand sanitizer).
3. Keeping things clean in my home (eg, phone, refrigerator, doorknobs).
4. Staying home as much as possible.
5. Wearing a mask or face covering when I go out of the house.
6. Staying at least six feet (about 3 steps) away from people I don't live with.
7. Avoiding gatherings or groups of other people.

We computed a mean of the response to these items to create a protective behavior adherence score (range 1-5). The internal consistency of this positive protective behaviors index was 0.85, and this was not improved by the deletion of any specific item.

Demographic Characteristics

Information on the demographics of the survey respondents, including age, gender, race or ethnicity, level of education, income, and political party affiliation, was obtained. Gender was initially assessed using 5 categories: male, female, transgender (identify as male), transgender (identify as female), and other. The responses were then collapsed into 2 categories ("identify as male" and "identify as female"). Race or ethnicity was coded as White, Black, Hispanic, multiracial, and other (which included American Indian, Asian, and other). Education was initially assessed with 10 strata, which were collapsed into 4 categories: none through high school or general education diploma, postsecondary (eg, trade school, some college, or associates), bachelor's, and advanced degree (eg, masters, doctoral or professional). Income was assessed with 9 strata, ranging from less than US $20,000 to more than US $150,000.

Survey Administration

The full survey assessed a range of individual and household characteristics, attitudes, and behaviors related to the COVID-19 pandemic. Surveys were completed through the Qualtrics web-based platform using a sample provided by Dynata [37]. Dynata's research panel comprises an opt-in list of over 60 million individuals globally. For this study we requested a nationally representative sample of 1000 US adults aged 18 years and above. Quotas were used to approximate national rates for age, gender, race, income, and US region. The survey was conducted as open enrollment, whereby eligible panel members who log into the Dynata website were offered a chance to participate in this survey. Participants received modest compensation (approximately US $1) from Dynata for completing the survey.

During the last week of May 2020, a total of 2272 individuals clicked on our survey invitation link, of which 187 did not complete an age screener item or consent, and 609 were ineligible for the survey or refused consent. This yielded 1476 complete survey responses from age-eligible, consenting individuals. To ensure the quality of the respondent data, we further excluded 402 survey responses based on either of two criteria. First, we excluded 375 survey responses from individuals who completed the entire survey in less than 10 minutes (the minimum time we considered needed to complete a valid survey). The mean time for survey completion for these excluded respondents was 5.4 (SD 3.3) minutes. Second, we excluded 27 survey responses from individuals who answered all items within a 16-item block of items assessing attitudes (and perceived norms) toward the pandemic with an identical response. This is the equivalent of clicking down an entire column (eg, all "Strongly Agree" or "Disagree" responses) for all items. Because some of the 16 items in this section were worded in the positive direction (eg, *Social distancing has slowed the spread of the coronavirus*) and the others, in the negative direction (eg, *Social distancing is not really doing much good*), we considered these "response set" patterns contradictory and a sign of poor-quality survey responses. Thus, we finally considered 1074 surveys for the present analyses.

The mean time to complete the survey was 25.3 (range 10.1-117.1) minutes for the included participants.

Hypotheses

We have two sets of hypotheses regarding the relationship between CoV-eHEALS scores and the participants' demographic characteristics and COVID-19–related KAPs.

Hypothesis 1

We expect to find significant associations between CoV-eHEALS scores and demographic characteristics. Specifically, we have the following expectations:

- Hypothesis 1a: CoV-eHEALS score will be negatively associated with age (ie, it will be lower among older individuals). This is based on several prior studies that found lower general eHEALS scores among older individuals [26,27,29,30].
- Hypothesis 1b: CoV-eHEALS score will be lower among ethnic minority groups. This is based on prior studies that have reported lower engagement with health information available on the internet or lower general eHEALS scores among minority populations [17,19,27,30].
- Hypothesis 1c: CoV-eHEALS score will be positively associated with educational attainment (ie, it will be higher among those who report completing higher formal education). This is based on several prior studies that identified this relationship between educational attainment and the general eHEALS measure [23,26,27,29,30].

Hypothesis 2

We expect to find significant associations between CoV-eHEALS scores and COVID-19 KAPs. Specifically, we have the following expectations:

- Hypothesis 2a: CoV-eHEALS score will be positively associated with COVID-19–related knowledge. This is based on prior studies showing a positive association between the general eHEALS measure and disease-specific knowledge or perceived understanding and knowledge of personal health issues [29,31].
- Hypothesis 2b: CoV-eHEALS score will be negatively associated with conspiracy beliefs. Mistrust of traditional information sources (eg, government, public health agencies, and mainstream media) is a core characteristic of individuals who hold conspiracy beliefs. We believe there is likely a negative association between a person's trust in these information sources and their confidence that they can find, assess, and use health information available on the internet.
- Hypothesis 2c: CoV-eHEALS score will be positively associated with adherence to behaviors that protect from COVID-19. This is based on prior studies showing more positive health behaviors (eg, healthy lifestyle and engagement in cancer screening) among individuals with higher general eHEALS scores [32,38].

Statistical Procedures

For hypothesis 1, we examined the relationship between demographic variables and the CoV-eHEALS score. Age and income (represented as 9 income strata) were examined as continuous variables. Gender, ethnicity, and educational attainment were examined as categorical variables. We first examined these associations separately using Pearson's correlation to examine the association between CoV-eHEALS score and continuous variables (eg, age and income) and analysis of variance to examine the association between CoV-eHEALS score and categorical variables (eg, gender, ethnicity, and educational attainment). We then examined the independent association between demographic variables and CoV-eHEALS score by using a linear regression model with CoV-eHEALS score as the dependent variable and demographic characteristics as the independent variables (with dummy coding of gender, ethnicity, and education).

For hypothesis 2, we examined the association between CoV-eHEALS score and COVID-19 knowledge, conspiracy beliefs, and protective behaviors. We performed multivariate analysis of variance (MANOVA) to simultaneously assess the relationship between these three dependent variables (scores for knowledge, conspiracy belief, and protective behavior adherence) and our main variable of interest (ie, low vs high CoV-eHEALS scores), while controlling for demographic characteristics as covariates (with age and income as continuous variables and dummy coding for gender, ethnicity, and education). To further illustrate the relationship between CoV-eHEALS scores and COVID-19 KAPs, we created simplified composite variables to represent each KAP measures. For knowledge, we computed a sum of the total number of knowledge items answered correctly (ie, answered "Definitely true" or "Probably true" for knowledge items 1-3 and "Definitely false" or "Probably false" for knowledge items 4-7) by each respondent (range 0-7). For conspiracy beliefs, we computed a sum of the total number of conspiracy items rejected (ie, answered "Definitely false" or "Probably false") by each respondent (range 0-3). For protective behaviors, we computed a sum of the total number of behaviors for which the respondent reported routine engagement (eg, answered "Always" or "Almost Always"). We then compared the distribution of these compositive variables for respondents classified as having low versus high CoV-eHEALS scores by using chi-square tests to assess statistical significance.

All analyses for this study were performed using SPSS software (version 25; IBM Corp).

Ethical Review

This survey study was reviewed and judged to be exempt (survey without identifying information) by the University of Michigan's institutional review board.

Results

The demographic characteristics of the study participants and their CoV-eHEALS and COVID-19 KAP scores are shown in Table 1. The sample comprised 55.6% (575/1074) female, 69.9% (723/1074) White, 8.1% (84/1074) Black, 9.2% (95/1074) Hispanic, and 6.3% (65/1074) multiracial participants. Their mean age was 47.3 (SD 17.1) years. Their mean CoV-eHEALS score was 29.0 (SD 6.1), and their mean scores for the COVID-19 KAP measures were as follows: knowledge 3.8 (SD

0.8), conspiracy beliefs 2.9 (SD 1.1), and protective behaviors index 3.9 (SD 0.9).

Results of the analyses related to hypothesis 1 (ie, Associations between CoV-eHEALS score and demographic characteristics) are shown in Tables 2 and 3. Using bivariate comparison, we found a significant association between the CoV-eHEALS score and income and educational attainment (Table 2). The correlation between CoV-eHEALS score and income was positive (r=0.087, P=.005) indicating individuals with higher income have higher coronavirus-related eHealth literacy. The direction of the relationship between CoV-eHEALS score and education was similar—individuals with higher educational attainment reported higher CoV-eHEALS scores. Results from a multivariate liner regression model are shown in Table 3. In this model, there are independent associations between CoV-eHEALS scores and ethnicity and educational attainment. The CoV-eHEALS score was lower for Black participants (standardized β=–.083, P=.02) than for White participants (the reference group). The CoV-eHEALS score was lower among participants who completed education only up to a high-school degree (standardized β=–.151, P=.001) than among those with advanced degrees (the reference group). In the regression model, the association with income was no longer significant. The association between CoV-eHEALS score and age was not significant in either the bivariate (r=0.009, P=.79) or multivariate (standardized β=–.038, P=.29) analysis.

Table 1. Characteristics of study participants (N=1074) and their mean scores for various study measures.

Variable	Value, n (%)
Age (years), n (%)	
18-35	304 (29.5)
36-50	263 (25.6)
51-65	277 (26.9)
≥65	185 (18)
Income (US $) , n (%)	
<30,000	291 (28.1)
30,000-74,999	397 (38.4)
≥75,000	346 (33.5)
Gender , n (%)	
Male	459 (44.4)
Female	575 (55.6)
Race or ethnicity, n (%)	
White	723 (69.9)
Black	84 (8.1)
Hispanic	95 (9.2)
Multiracial	65 (6.3)
Other	67 (6.5)
Education, n (%)	
Up to high school or GED[a]	225 (21.8)
Postsecondary (eg, trade school, some college, or associates)	326 (31.6)
Bachelor's degree	310 (30)
Advanced degree (eg, Masters, Doctoral or Professional)	172 (16.7)
Scores, mean (SD)	
Coronavirus-related eHealth Literacy Scale (range 8-40)	29.0 (6.1)
Knowledge (range 1-5)	3.8 (0.8)
Conspiracy beliefs (range 1-5)	2.9 (1.1)
Positive behavior adherence (range 1-5)	3.9 (0.9)

[a]GED: Tests of General Educational Development.

Table 2. Bivariate association between demographic characteristics and coronavirus-related eHealth literacy.

Variable	eHealth literacy score, mean (SD)	*P* value
Gender		.47
Male	29.2 (6.3)	
Female	28.9 (5.9)	
Ethnicity		.21
White	29.1 (6.0)	
Black	27.6 (5.7)	
Multiracial	28.9 (6.3)	
Hispanic	29.4 (5.9)	
Other	29.3 (7.1)	
Education		<.001
Up to high school or GED[a]	27.6 (6.6)	
Postsecondary (eg, trade school, some college, or associates)	28.8 (6.0)	
Bachelor's degree	29.9 (5.7)	
Advanced degree (eg, Masters, Doctoral or Professional)	30.0 (5.5)	

[a]GED: Tests of General Educational Development.

Table 3. Independent association between demographic characteristics and coronavirus-related eHealth literacy.

Variable	Standardized β coefficient	*P* value
Age (continuous)	–.038	.29
Income (continuous, 9 strata)	.023	.21
Gender		
Male	Ref[a]	
Female	–.003	.92
Ethnicity		
White	Ref	
Black	–.083	*.02* [b]
Multiracial	–.018	.60
Hispanic	–.006	.86
Other	–.014	.68
Education		
Up to high school or GED[c]	–.151	*.001* [b]
Postsecondary (eg, trade school, some college, or associates)	–.079	.09
Bachelor's	–.007	.87
Advanced degree (eg, Masters, Doctoral or Professional)	Ref	

[a]Ref: reference value.

[b]Italicized values indicate statistical significance.

[c]GED: Tests of General Educational Development.

Results of the analyses related to hypothesis 2 (ie, association between CoV-eHEALS and COVID-19 KAP scores) are shown in Table 4. Among our respondents, 29% (306/1074) were classified as having low coronavirus-related eHealth literacy (ie, total CoV-eHEALS score <26). Results from the MANOVA model show a significant association between CoV-eHEALS score and all three (COVID-19 knowledge, conspiracy beliefs, and adherence to protective behaviors) dependent variables ($F_{3,1013}$=20.89, P<.001; Wilks Λ=0.94, partial η^2=0.058). When adjusting for sociodemographic differences, respondents with higher (vs lower) CoV-eHEALS scores had higher mean scores

for knowledge and protective behaviors adherence and a lower mean score for conspiracy beliefs.

The nature of the relationship between CoV-eHEALS and COVID-19 KAP scores is further illustrated in Figures 1-3. Figure 1 compares the number of knowledge items answered correctly by respondents with low versus high CoV-eHEALS scores. For those with low CoV-eHEALS scores, nearly 33% (99/305) correctly answered 2 or fewer knowledge items, and less than 10% (27/305) correctly answered all 7 items. For those with high CoV-eHEALS scores, less than 20% (88/742) correctly answered 2 or fewer knowledge items and nearly 25% (175/742) correctly answered all 7 items. The difference in the distribution of the number of correct answers on these 7 knowledge items was statistically significant (P<.001).

Figure 2 compares the number of rejected conspiracy items (ie, participants who assessed the conspiracy item as "Definitely False" or "Probably False") for participants with low versus high CoV-eHEALS scores. For those with low CoV-eHEALS scores, approximately 50% (153/305) did not reject any of these items, whereas only 8% (25/305) of them rejected all 3 items. For those with high CoV-eHEALS scores, 35% (261/742) did not reject any of these items, whereas 21% (158/742) rejected all 3 items. The difference in the distribution of the number of rejected conspiracy items was statistically significant (P<.001).

Figure 3 compares the number of protective behaviors routinely reported by the participants (ie, "Always" or "Almost Always" engaging in this practice) with low versus high CoV-eHEALS scores. For those with low CoV-eHEALS scores, nearly 30% (85/305) reported engaging in 2 or fewer behaviors, and a similar proportion of the participants (107/305, 35%) reported routine practice of 6 or all 7 of the protective behaviors. For those with high CoV-eHEALS scores, about 15% (109/742) of the participants reported engaging in 2 or fewer behaviors, whereas just over 50% (487/742) reported routine practice of 6 or all 7 of the protective behaviors. The difference in the distribution of the number of routine protective behaviors was statistically significant (P<.001).

Table 4. COVID-19 knowledge, conspiracy beliefs, and protective behaviors for respondents with low and high coronavirus-related eHealth literacy.

CoV-eHEALS[a] score	Estimated mean score[b] (SE)		
	Knowledge	Conspiracy beliefs	Protective behaviors
Low score (n=298)	3.6 (0.040)	3.0 (0.064)	3.6 (0.049)
High score (n=729)	3.9 (0.025)	2.8 (0.040)	4.0 (0.031)
P value	<.001	.03	<.001

[a]CoV-eHEALS: coronavirus-related eHealth literacy scale.

[b]Estimated means adjusted for age, income, gender, ethnicity, and education level. Overall multivariate analysis of variance model; Box M=53.35; F=8.86; P<.001.

Figure 1. Number of correct knowledge items by coronavirus-related eHealth literacy.

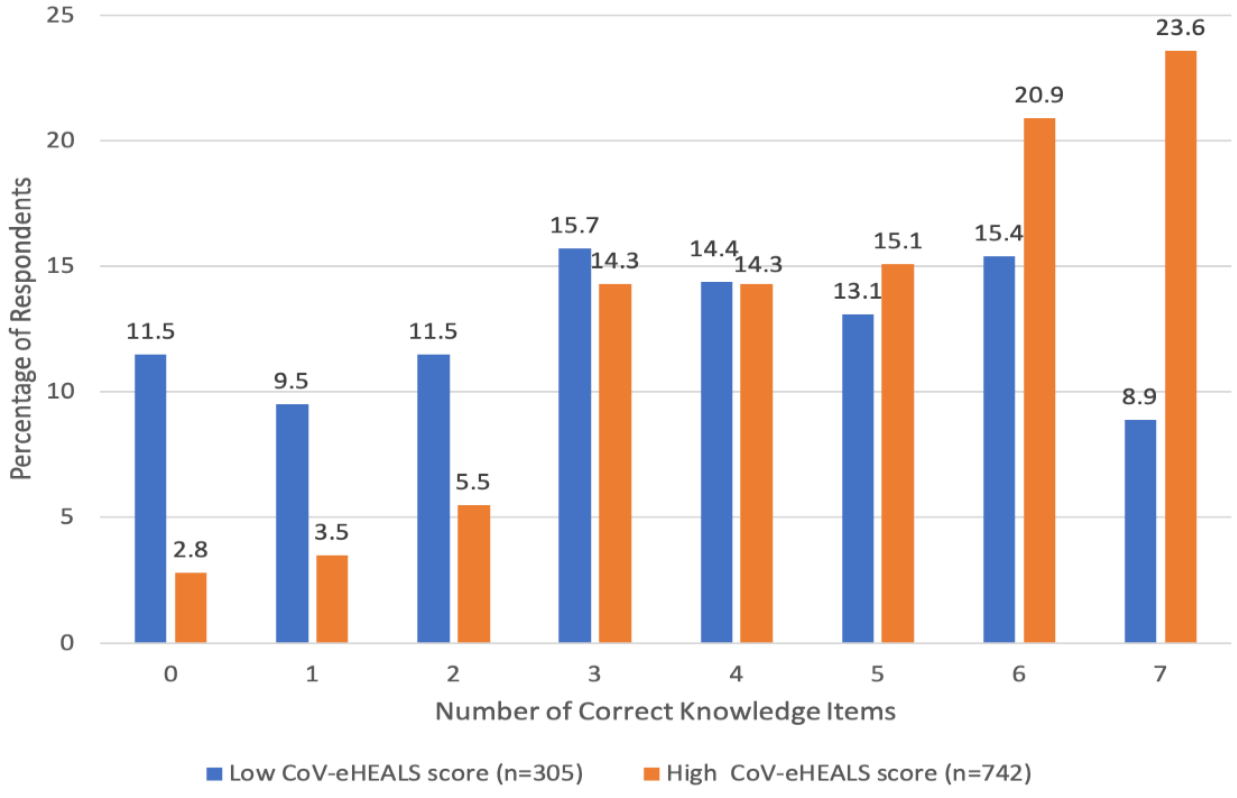


Figure 2. Number of rejected conspiracy items by coronavirus-related eHealth literacy.

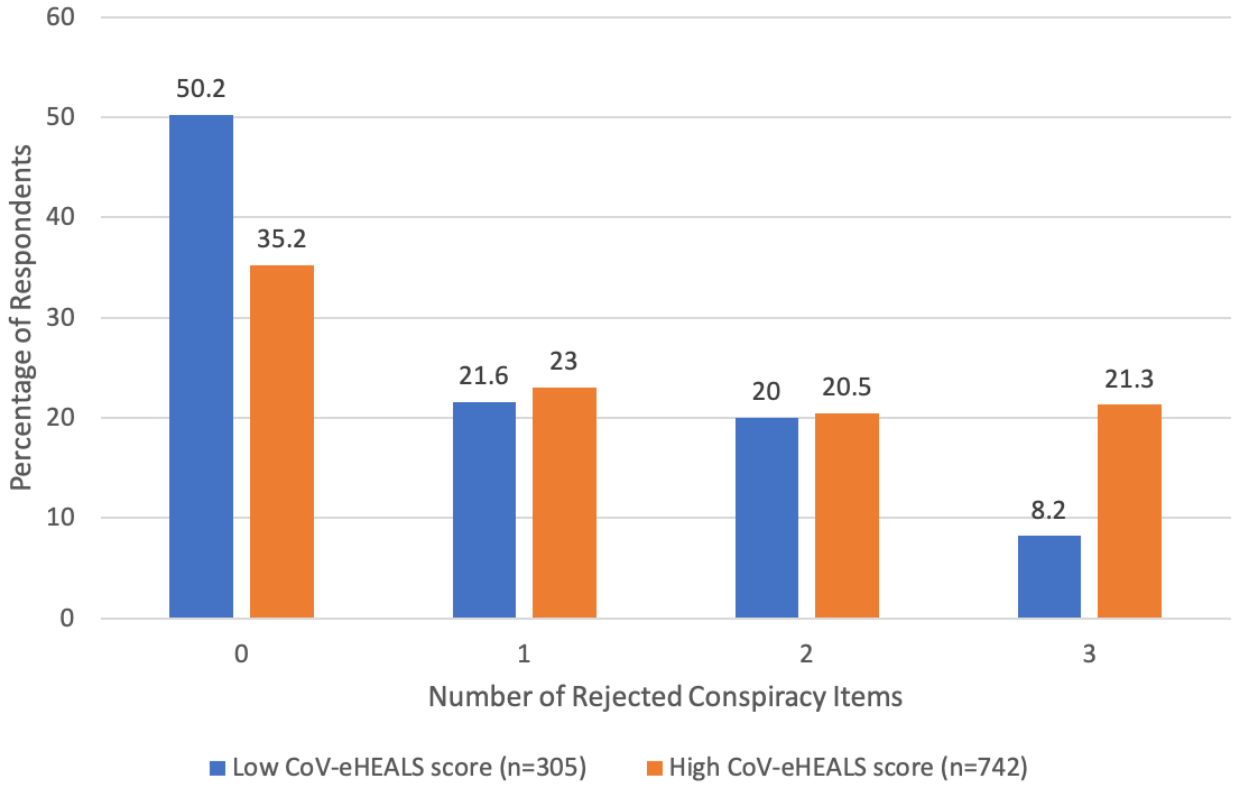
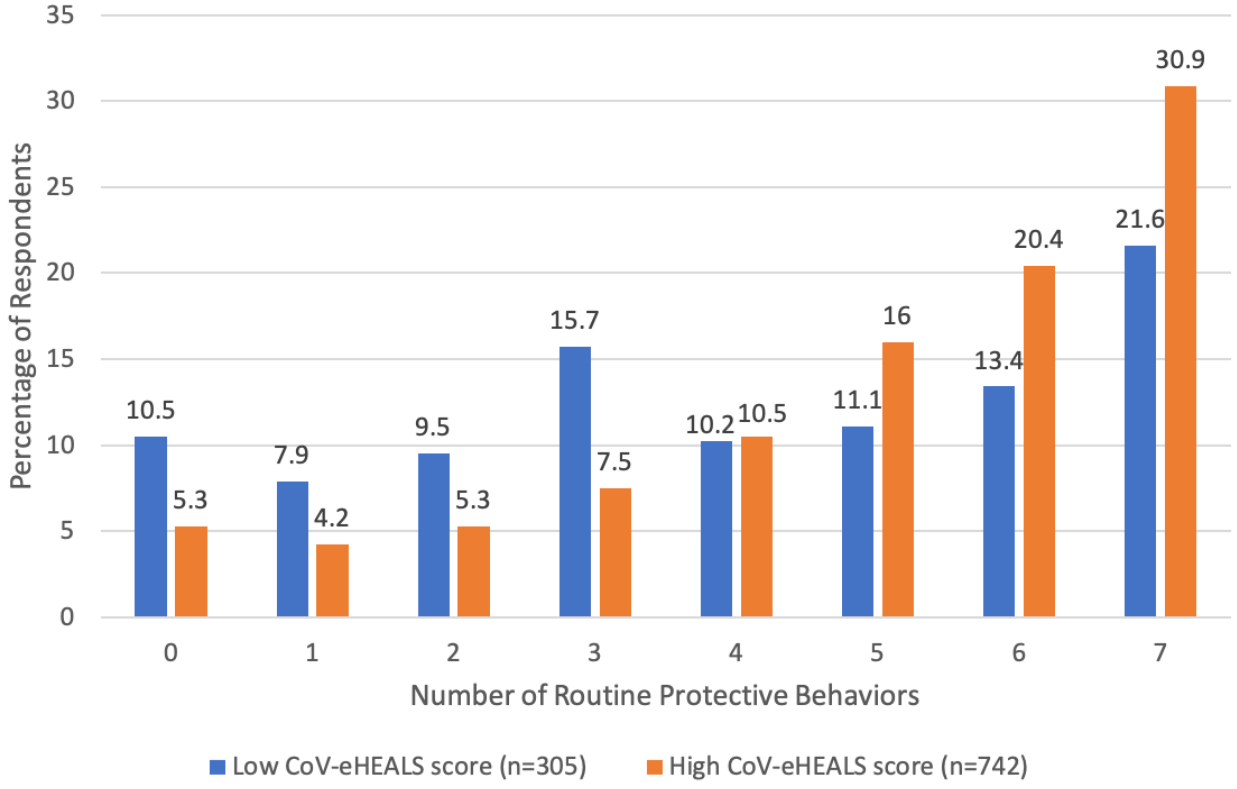


Figure 3. Number of routine protective behaviors by coronavirus-related eHealth literacy.



Discussion

The principal findings of the study show a clear and consistent association between higher coronavirus-related eHealth literacy and greater knowledge, lower conspiracy beliefs, and greater engagement in protective behaviors. The mean CoV-eHEALS score used in this study was similar to those used for the general eHEALS in several population samples [23,26,27]; it was

somewhat higher than that reported among some disease-specific groups (eg, cardiovascular disease, lung cancer, patients with HIV) [22,24,39]. It is possible that widespread public attention, media coverage (including content available on the internet), and concerted effort of major public health organizations to disseminate health information on the internet (eg, World Health Organization and the Centers for Disease Control) during the current pandemic contributed to this finding.

It is important to acknowledge that we administered a modified version of the eHEALS measure that was specific to information about coronavirus (ie, CoV-eHEALS). Although we recognize that this is not typical practice for evaluation of eHealth literacy, we believe this was appropriate given the critical need to assess and understand the ability of individuals and the public to find, assess, and use information available on the internet that is specific to the coronavirus during the current COVID-19 pandemic. The findings we report for this CoV-eHEALS measure are consistent with those recently reported by other teams that administered the general eHEALS measure as part of pandemic-related studies. For instance, a study by Do et al [40] of health care workers in Vietnam reported a significant positive association between participants' general eHEALS score and their self-reported adherence to occupational infection prevention and control measures. In a national web-based survey of internet users in China, Li and Liu [41] found a significant association between the general eHEALS measure and self-reported practice of protective behaviors against COVID-19 [41]. Future work could examine the relation between CoV-eHEALS and general eHEALS measures and COVID-19−related KAPs.

The results of this study largely support our first set of hypotheses regarding the association between CoV-eHEALS scores and demographic characteristics. Multivariate analyses showed that Black participants had lower CoV-eHEALS scores than White participants (hypothesis 1b). This finding is consistent with some prior studies that have found low general eHEALS scores and a low frequency of seeking health information on the internet among ethnic minority groups [15,17,19,27]. The finding of lower CoV-eHEALS scores among Black participants is of particular concern given the recognized disparities in the impact of COVID-19 on minority groups. Moreover, our finding of lower CoV-eHEALS scores among those with lower educational attainment (hypothesis 1c) is consistent with the same association observed in multiple previous studies using the general eHEALS measure [23,26,27,29,30]. When considered along with the finding of lower CoV-eHEALS scores among Black respondents, this finding reinforces the continued need to address COVID-19−related health disparities that place an undue burden on underserved and disadvantaged groups. Contrary to our original hypothesis 1a, we did not find an association between the participants' CoV-eHEALS scores and age. This finding is in contrast with a number of prior studies that have identified more advanced age as a predictor of lower general eHEALS scores. The lack of a decrease in CoV-eHEALS score among older participants is somewhat reassuring, particularly given that older adults are at a greater risk of serious illness or death due to COVID-19.

Our study findings also consistently support our second set of hypotheses regarding the association between coronavirus-related eHealth literacy and COVID-19 KAPs. Our analyses showed a significant association in the expected directions between the CoV-eHEALS measure and COVID-19 knowledge, conspiracy beliefs, and engagement in protective behaviors. In considering these findings, it is important to recognize that self-efficacy is a central concept underlying the development of the general eHEALS and, consequently, also for this adapted CoV-eHEALS measure. Although the assessment of self-efficacy is a critical aspect of many major theories of health behavior, it is also recognized that individuals commonly overestimate their abilities to perform more complex tasks [42-45]. For the general eHEALS measure, studies examining the association between self-reported eHealth literacy and the actual performance of functional measures of internet skills have reported variable results. Van Der Vaart and colleagues [28] noted that, in a small sample of patients with rheumatic disease, an individual's eHEALS score was not consistently related to their actual performance of internet skills. In contrast, Neter and Brainin [46] found a positive, if modest, association between the self-reported general eHEALS measure and observed performance on health-related internet tasks. The nature of the relationship observed between our CoV-eHEALS measure and knowledge provide further insights into this issue. The overall positive and independent association between the CoV-eHEALS and knowledge scores lends support to the validity of the eHEALS approach. At the same time, specific findings also suggest that individuals may overestimate their abilities to some degree. Among those with high CoV-eHEALS scores (ie, high confidence in their own ability to find, assess, and use information about coronavirus on the internet), only about 1 in 4 participants was able to correctly answer all 7 items on our COVID-19 knowledge scale.

The relationship observed between the CoV-eHEALS measure and conspiracy beliefs also warrants further discussion. For our study participants, the mean score on the conspiracy beliefs scale was 2.9, which indicates that, on average, our sample was "unsure" about the truth or falsehood of these statements. Although differences in the wording of questions and format preclude direct comparisons, other studies have reported high rates of endorsement of COVID-19 conspiracy beliefs [47,48]. Our finding of greater rejection of conspiracy beliefs among those with higher CoV-eHEALS scores is consistent with the work of Richtering and colleagues [22] who found a positive association between the eHEALS score and specific facets of a general health literacy measure, including the ability to perform "critical appraisal" of information available on the internet. Although the negative association between CoV-eHEALS and conspiracy scale scores observed in our study is encouraging, it is important to recognize that even among respondents who could be considered as having higher CoV-eHEALS scores (ie, total score ≥26), fewer than half of the participants clearly rejected two or more of the conspiracy items. One study, involving a national survey in Poland, actually reported greater acceptance of conspiracy beliefs among individuals with higher general eHealth literacy [49]. In considering the relationship between CoV-eHEALS scores or general eHEALS scores and these beliefs, it is important to

recognize the complex interplay of factors that influence conspiracy thinking such as underlying social orientation (eg, individualist vs collectivist), perception of power or powerlessness, ideology or political affiliation, and media consumption [48,50]. All these factors highlight the importance of considering a broader framework for managing the COVID-19 infodemic. Eysenbach [4] recently presented an "information cake" model that includes building eHealth literacy (and general scientific literacy) along with information monitoring, encouraging knowledge refinement and information quality management, and accurate and timely knowledge translation as the four pillars of effective infodemic management.

In the body of published work on the eHEALS measure, relatively few studies have reported on the relationship between eHEALS scores and specific health behaviors or health outcomes. Neter and colleagues [51] recently reviewed this topic and concluded that although there are some positive associations, additional study on this topic is needed. In the Neter review [51], the most consistent associations seemed to be between eHEALS and health behaviors. For example, Mitsutake and colleagues reported in separate studies the finding of positive association between scores on the general eHEALS measure and health-promoting behaviors (eg, physical activity, healthy eating) and also colorectal cancer screening practice [32,38]. We found that compared to individuals with low CoV-eHEALS scores, those with high CoV-eHEALS scores reported engaging in routine practice of one additional protective behavior (eg, mask wearing, social distancing). These observed relationships are both statistically significant and meaningful from a personal and public health standpoint. Our finding of a positive association between the CoV-eHEALS score and engagement in protective behaviors contributes to our understanding of the eHealth literacy and provides further evidence of the relationship between measures of eHealth literacy (eg, general eHEALS or our CoV-eHEALS) and actual health behaviors and practices.

There are several limitations to consider when interpreting our study findings. First, the results reported here are from a single cross-sectional survey, and thus, we cannot make claims regarding causation. For example, although we did find a negative association between CoV-eHEALS scores and conspiracy beliefs, we cannot be certain whether a higher CoV-eHEALS score led to reduced acceptance of these beliefs or whether a predisposition to conspiracy thinking led to lower CoV-eHEALS scores. Second, it is important to acknowledge that the CoV-eHEALS and COVID-19 KAP measures are based on self-report. The need for further study of the relationship between self-reported eHEALS measures and actual performance has been discussed above. Associations between trajectories of self-reported protective behaviors and COVID-19 cases supports the validity of these self-report measures; however, the precise relationship between self-reported and actual behavior (eg, difference between behavior that is reported to occur "some of the time" vs "almost all of the time") requires additional study [52]. Third, this survey was conducted during a single, brief time period in one specific country relatively soon after the onset of the COVID-19 pandemic. Further work will be needed to determine how CoV-eHEALS scores may vary across different countries and how this might change over time. Fourth, it is important to note that this survey was performed using a web-based survey format. Although the rates of internet access in the United States are quite high in general and characteristics of our study sample were similar to that of the general US population, the administration of a web-based survey could certainly bias the sample toward individuals with greater familiarity with technology and the internet. Computer and internet use are well-recognized predictors of eHEALS, which we did not specifically assess in this study [22,23,27,30]. As such, our estimate of 29% of the US adult population having low CoV-eHEALS scores should likely be considered a lower bound for this estimate.

Despite these limitations, there are some potentially important implications related to our study findings. We found that although the overall level of coronavirus-related eHealth literacy in this study was relatively high, there still remains a substantial proportion of the US adult population that has low coronavirus-related eHealth literacy; this population might thus be considered at higher risk of negative COVID-19 KAPs. Recent studies assessing the quality of health information available on the internet about COVID-19 have found inconsistent coverage of key public health recommendations with a majority of websites having moderate-to-low quality scores [53,54]. These authors identified substantial opportunities to improve the clarity of presentation of critical health information and argued that broader implementation and adherence to quality standards for presentation of COVID-19−related information available on the internet could be helpful in terms of improving public health literacy on this topic. It is important to note that some major search engines have taken specific steps to improve search simplicity and delivery of high-quality COVID-19−related health information. For example, Google has a dedicated landing page for COVID-19−related information that is displayed following general searches about coronavirus or COVID-19. This landing page highlights key topics (eg, disease trends; access to testing; and health information on symptoms, prevention, and treatments) with summaries of key facts and direct links leading to high-quality information sources. Finally, it is important to recognize that searching for information on the internet can be a complicated and challenging process [55]. Despite the capabilities of modern search engines, people are more effective in finding accurate information if they possess some basic skills in how to find and use information on the web (eg, knowing to "click through" to view the actual website instead of relying upon websites summaries, checking timeliness and quality of information sources, and cross-checking different information sources) [56]. Deficits in such search skills are common in the general population; nevertheless, prior work shows that through web-based training, people can develop these skills and improve their abilities to find high-quality and accurate information on the internet.

Given the consistent associations between CoV-eHEALS scores and COVID-19 KAPs, there may be some benefit in teaching such search skills in general or specifically in terms of searches for COVID-19−related information. In the future, it could be

important to assess and track coronavirus-related eHealth literacy at the individual and population levels. Identifying and addressing low coronavirus-related eHealth literacy could prove helpful in improving COVID-19−related knowledge, attitudes, and practices, thereby reducing future illness and deaths during this pandemic.

Acknowledgments

This work was supported by National Cancer Institute Grant P30CA046592-29-S4 and a Google Focus award (to LA).

Conflicts of Interest

None declared.

References

1. Brørs G, Norman C, Norekvål TM. Accelerated importance of eHealth literacy in the COVID-19 outbreak and beyond. Eur J Cardiovasc Nurs 2020 Aug;19(6):458-461 [FREE Full text] [doi: 10.1177/1474515120941307] [Medline: 32667217]
2. Chong Y, Cheng H, Chan H, Chien W, Wong S. COVID-19 pandemic, infodemic and the role of eHealth literacy. Int J Nurs Stud 2020 Aug;108:103644 [FREE Full text] [doi: 10.1016/j.ijnurstu.2020.103644] [Medline: 32447127]
3. Sentell T, Vamos S, Okan O. Interdisciplinary perspectives on health literacy research around the world: more important than ever in a time of COVID-19. Int J Environ Res Public Health 2020 Apr 26;17(9):3010 [FREE Full text] [doi: 10.3390/ijerph17093010] [Medline: 32357457]
4. Eysenbach G. How to fight an infodemic: the four pillars of infodemic management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
5. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-10-06]
6. Coronavirus Disease 2019 (COVID-19). Center for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/index.html [accessed 2020-10-06]
7. Clements JM. Knowledge and behaviors toward COVID-19 among US residents during the early days of the pandemic: cross-sectional online questionnaire. JMIR Public Health Surveill 2020 May 08;6(2):e19161 [FREE Full text] [doi: 10.2196/19161] [Medline: 32369759]
8. McFadden S, Malik A, Aguolu O, Willebrand K, Omer S. Perceptions of the adult US population regarding the novel coronavirus outbreak. PLoS One 2020;15(4):e0231808 [FREE Full text] [doi: 10.1371/journal.pone.0231808] [Medline: 32302370]
9. Alobuia W, Dalva-Baird N, Forrester J, Bendavid E, Bhattacharya J, Kebebew E. Racial disparities in knowledge, attitudes and practices related to COVID-19 in the USA. J Public Health (Oxf) 2020 Aug 18;42(3):470-478 [FREE Full text] [doi: 10.1093/pubmed/fdaa069] [Medline: 32490519]
10. Block R, Berg A, Lennon RP, Miller EL, Nunez-Smith M. African American adherence to COVID-19 public health recommendations. Health Lit Res Pract 2020 Aug 06;4(3):e166-e170 [FREE Full text] [doi: 10.3928/24748307-20200707-01] [Medline: 32926172]
11. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 06;10(1):16598 [FREE Full text] [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]
12. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for managing the COVID-19 infodemic: methods and results of an online, crowdsourced WHO technical consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
13. Frenkel S, Decker B, Alba D. How the 'Plandemic' Movie and Its Falsehoods Spread Widely Online. The New York Times. URL: https://www.nytimes.com/2020/05/20/technology/plandemic-movie-youtube-facebook-coronavirus.html
14. Freeman D, Waite F, Rosebrock L, Petit A, Causier C, East A, et al. Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England. Psychol Med 2020 May 21:1-13 [FREE Full text] [doi: 10.1017/S0033291720001890] [Medline: 32436485]
15. Rice R. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
16. Powell J, Inglis N, Ronnie J, Large S. The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. J Med Internet Res 2011 Feb 23;13(1):e20 [FREE Full text] [doi: 10.2196/jmir.1600] [Medline: 21345783]
17. Bangerter L, Griffin J, Harden K, Rutten L. Health information-seeking behaviors of family caregivers: analysis of the Health Information National Trends Survey. JMIR Aging 2019 Jan 14;2(1):e11237 [FREE Full text] [doi: 10.2196/11237] [Medline: 31518309]

18. Kontos E, Blake K, Chou W, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
19. Sherman L, Patterson M, Tomar A, Wigfall L. Use of digital health information for health information seeking among men living with chronic disease: data from the Health Information National Trends Survey. Am J Mens Health 2020;14(1):1557988320901377 [FREE Full text] [doi: 10.1177/1557988320901377] [Medline: 31973642]
20. Norman C, Skinner H. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
21. Paige S, Krieger J, Stellefson M, Alber J. eHealth literacy in chronic disease patients: An item response theory analysis of the eHealth literacy scale (eHEALS). Patient Educ Couns 2017 Feb;100(2):320-326 [FREE Full text] [doi: 10.1016/j.pec.2016.09.008] [Medline: 27658660]
22. Richtering S, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth literacy: predictors in a population with moderate-to-high cardiovascular risk. JMIR Hum Factors 2017 Jan 27;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
23. Chung S, Nahm E. Testing reliability and validity of the eHealth Literacy Scale (eHEALS) for older adults recruited online. Comput Inform Nurs 2015 Apr;33(4):150-156 [FREE Full text] [doi: 10.1097/CIN.0000000000000146] [Medline: 25783223]
24. Milne RA, Puts MTE, Papadakos J, Le LW, Milne VC, Hope AJ, et al. Predictors of high eHealth literacy in primary lung cancer survivors. J Cancer Educ 2015 Dec;30(4):685-692. [doi: 10.1007/s13187-014-0744-5] [Medline: 25355524]
25. Nguyen J, Moorhouse M, Curbow B, Christie J, Walsh-Childers K, Islam S. Construct validity of the eHealth Literacy Scale (eHEALS) among two adult populations: a Rasch analysis. JMIR Public Health Surveill 2016 May 20;2(1):e24 [FREE Full text] [doi: 10.2196/publichealth.4967] [Medline: 27244771]
26. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
27. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
28. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
29. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
30. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: A systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
31. Stellefson ML, Shuster JJ, Chaney BH, Paige SR, Alber JM, Chaney JD, et al. Web-based health information seeking and eHealth literacy among patients living with chronic obstructive pulmonary disease (COPD). Health Commun 2018 Dec;33(12):1410-1424 [FREE Full text] [doi: 10.1080/10410236.2017.1353868] [Medline: 28872905]
32. Mitsutake S, Shibata A, Ishii K, Oka K. Association of eHealth literacy with colorectal cancer knowledge and screening practice among internet users in Japan. J Med Internet Res 2012 Nov 13;14(6):e153 [FREE Full text] [doi: 10.2196/jmir.1927] [Medline: 23149453]
33. Li S, Feng B, Liao W, Pan W. Internet use, risk awareness, and demographic characteristics associated with engagement in preventive behaviors and testing: cross-sectional survey on COVID-19 in the United States. J Med Internet Res 2020 Jun 16;22(6):e19782 [FREE Full text] [doi: 10.2196/19782] [Medline: 32501801]
34. van Prooijen JW, van Vugt M. Conspiracy theories: evolved functions and psychological mechanisms. Perspect Psychol Sci 2018 Nov;13(6):770-788 [FREE Full text] [doi: 10.1177/1745691618774270] [Medline: 30231213]
35. Centers for Disease Control and Prevention - COVID-19. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html [accessed 2021-02-24]
36. Toussaint L, Cheadle A, Fox J, Williams D. Clean and contain: initial development of a measure of infection prevention behaviors during the COVID-19 pandemic. Ann Behav Med 2020 Sep 01;54(9):619-625 [FREE Full text] [doi: 10.1093/abm/kaaa064] [Medline: 32856691]
37. Dynata - The Data Platform for [richer] insights. URL: https://www.dynata.com [accessed 2021-02-25]
38. Mitsutake S, Shibata A, Ishii K, Oka K. Associations of eHealth literacy with health behavior among adult internet users. J Med Internet Res 2016 Jul 18;18(7):e192 [FREE Full text] [doi: 10.2196/jmir.5413] [Medline: 27432783]
39. Robinson C, Graham J. Perceived Internet health literacy of HIV-positive people through the provision of a computer and Internet health education intervention. Health Info Libr J 2010 Dec;27(4):295-303 [FREE Full text] [doi: 10.1111/j.1471-1842.2010.00898.x] [Medline: 21050372]
40. Do B, Tran T, Phan D, Nguyen HC, Nguyen TTP, Nguyen HC, et al. Health literacy, eHealth literacy, adherence to infection prevention and control procedures, lifestyle changes, and suspected COVID-19 symptoms among health care workers

during lockdown: online survey. J Med Internet Res 2020 Nov 12;22(11):e22894 [FREE Full text] [doi: 10.2196/22894] [Medline: 33122164]
41. Li X, Liu Q. Social media use, eHealth literacy, disease knowledge, and preventive behaviors in the COVID-19 pandemic: cross-sectional study on Chinese netizens. J Med Internet Res 2020 Oct 09;22(10):e19684 [FREE Full text] [doi: 10.2196/19684] [Medline: 33006940]
42. Bandura A. A social cognitive theory of personality. In: Pervin L, John O, editors. Handbook of personality (2nd ed.). New York, NY: Guilford Publications; 1999:154-196.
43. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
44. Moores T, Chang J. Self-efficacy, overconfidence, and the negative effect on subsequent performance: A field study. Information & Management 2009 Mar;46(2):69-76. [doi: 10.1016/j.im.2008.11.006]
45. Weinberg B. A model of overconfidence. Pacific Economic Review 2009 Nov;14(4):502-515. [doi: 10.1111/j.1468-0106.2009.00466.x]
46. Neter E, Brainin E. Perceived and performed eHealth literacy: survey and simulated performance test. JMIR Hum Factors 2017 Jan 17;4(1):e2 [FREE Full text] [doi: 10.2196/humanfactors.6523] [Medline: 28096068]
47. Earnshaw V, Eaton L, Kalichman S, Brousseau N, Hill E, Fox A. COVID-19 conspiracy beliefs, health behaviors, and policy support. Transl Behav Med 2020 Oct 08;10(4):850-856 [FREE Full text] [doi: 10.1093/tbm/ibaa090] [Medline: 32910819]
48. Romer D, Jamieson K. Conspiracy theories as barriers to controlling the spread of COVID-19 in the U.S. Soc Sci Med 2020 Oct;263:113356 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113356] [Medline: 32967786]
49. Duplaga M. The determinants of conspiracy beliefs related to the COVID-19 pandemic in a nationally representative sample of internet users. Int J Environ Res Public Health 2020 Oct 26;17(21):7818 [FREE Full text] [doi: 10.3390/ijerph17217818] [Medline: 33114542]
50. Biddlestone M, Green R, Douglas K. Cultural orientation, power, belief in conspiracy theories, and intentions to reduce the spread of COVID-19. Br J Soc Psychol 2020 Jul;59(3):663-673 [FREE Full text] [doi: 10.1111/bjso.12397] [Medline: 32592420]
51. Neter E, Brainin E. Association between health literacy, eHealth literacy, and health outcomes among patients with long-term conditions. European Psychologist 2019;24(1):68-81. [doi: 10.1027/1016-9040/a000350]
52. Lazer L, Santillana M, Perlis R, Quintana A, Ognyanova K, Green J, et al. A 50-State COVID-19 Survey: Report #26: Trajectory of COVID-19-Related Behaviors. The COVID States Project. 2020 Nov. URL: https://kateto.net/covid19/COVID19%20CONSORTIUM%20REPORT%2026%20BEHAVIORS%20NOV%202020.pdf [accessed 2021-02-24]
53. Hernández-García I, Giménez-Júlvez T. Assessment of health information about COVID-19 prevention on the internet: infodemiological study. JMIR Public Health Surveill 2020 Apr 01;6(2):e18717 [FREE Full text] [doi: 10.2196/18717] [Medline: 32217507]
54. Jayasinghe R, Ranasinghe S, Jayarajah U, Seneviratne S. Quality of online information for the general public on COVID-19. Patient Educ Couns 2020 Aug 07:2594-2597 [FREE Full text] [doi: 10.1016/j.pec.2020.08.001] [Medline: 32800422]
55. Russell D. The Joy of Search - A Google Insider's Guide to Going Beyond the Basics. Cambridge, MA: MIT Press; Sep 2019.
56. Russell D, Callegaro M. How to Be a Better Web Searcher: Secrets from Google Scientists. Scientific American. 2019 Mar 26. URL: https://blogs.scientificamerican.com/observations/how-to-be-a-better-web-searcher-secrets-from-google-scientists/ [accessed 2021-02-24]

Abbreviations

CoV-eHEALS: Coronavirus-Related eHealth Literacy Scale
eHEALS: eHealth Literacy Scale
KAP: knowledge, attitude, and practice
MANOVA: multivariate analysis of variance

Edited by G Eysenbach, C Basch; submitted 20.10.20; peer-reviewed by E Neter, A Dominguez-Rodriguez, D Flemming, M Fiordelli; comments to author 10.11.20; revised version received 17.12.20; accepted 19.02.21; published 29.03.21.

Please cite as:
An L, Bacon E, Hawley S, Yang P, Russell D, Huffman S, Resnicow K
Relationship Between Coronavirus-Related eHealth Literacy and COVID-19 Knowledge, Attitudes, and Practices among US Adults: Web-Based Survey Study
J Med Internet Res 2021;23(3):e25042
URL: https://www.jmir.org/2021/3/e25042
doi:10.2196/25042
PMID:33626015

Original Paper

Acceptance of Technologies for Aging in Place: A Conceptual Model

Christina Jaschinski[1], PhD; Somaya Ben Allouch[2], PhD; Oscar Peters[3], PhD; Ricardo Cachucho[4], PhD; Jan A G M van Dijk[5], PhD

[1]Research Group Technology, Health & Care, Saxion University of Applied Sciences, Enschede, Netherlands

[2]Digital Life, Amsterdam University of Applied Sciences, Amsterdam, Netherlands

[3]Saxion University of Applied Sciences, Enschede, Netherlands

[4]Leiden Institute of Advanced Computer Science (LIACS), Leiden University, Leiden, Netherlands

[5]University of Twente, Enschede, Netherlands

Corresponding Author:
Christina Jaschinski, PhD
Research Group Technology, Health & Care
Saxion University of Applied Sciences
M.H. Tromplaan 28
Enschede, 7513AB
Netherlands
Phone: 31 6 57 81 46 14
Email: c.jaschinski@saxion.nl

Abstract

Background: Older adults want to preserve their health and autonomy and stay in their own home environment for as long as possible. This is also of interest to policy makers who try to cope with growing staff shortages and increasing health care expenses. Ambient assisted living (AAL) technologies can support the desire for independence and aging in place. However, the implementation of these technologies is much slower than expected. This has been attributed to the lack of focus on user acceptance and user needs.

Objective: The aim of this study is to develop a theoretically grounded understanding of the acceptance of AAL technologies among older adults and to compare the relative importance of different acceptance factors.

Methods: A conceptual model of AAL acceptance was developed using the theory of planned behavior as a theoretical starting point. A web-based survey of 1296 older adults was conducted in the Netherlands to validate the theoretical model. Structural equation modeling was used to analyze the hypothesized relationships.

Results: Our conceptual model showed a good fit with the observed data (root mean square error of approximation 0.04; standardized root mean square residual 0.06; comparative fit index 0.93; Tucker-Lewis index 0.92) and explained 69% of the variance in intention to use. All but 2 of the hypothesized paths were significant at the $P<.001$ level. Overall, older adults were relatively open to the idea of using AAL technologies in the future (mean 3.34, SD 0.73).

Conclusions: This study contributes to a more user-centered and theoretically grounded discourse in AAL research. Understanding the underlying behavioral, normative, and control beliefs that contribute to the decision to use or reject AAL technologies helps developers to make informed design decisions based on users' needs and concerns. These insights on acceptance factors can be valuable for the broader field of eHealth development and implementation.

(J Med Internet Res 2021;23(3):e22613) doi:10.2196/22613

KEYWORDS

ambient assisted living; assistive technology; healthy aging; technology adoption; theory of planned behavior; structural equation modeling

Introduction

Background

Demographic predictions show a growing number of people at risk for age-related chronic diseases and with a potential need for long-term care. At the same time, there is a growing shortage of caregivers. With the pressing demand for care, the workload for formal and informal caregivers is steadily increasing, negatively affecting their physical and mental well-being [1,2]. These developments put the sustainability of our current health care system at risk [3].

To address these challenges, European care reforms have induced a shift from institutionalized care to more care at home and aging in place. Similarly, the European Union (EU) has embraced an active aging policy strategy that emphasizes good health, security, and participation [4,5]. State-of-the-art assistive technologies, also known as ambient assisted living (AAL) technologies, are viewed as a vital contributor to this strategy.

Ambient Assisted Living

The term AAL has been introduced by the EU to describe the use of a new generation of information and communication technology (ICT)-based assistive technologies that provide holistic support to older adults in managing their health, remaining independent, and staying involved with their community. AAL technologies are also directed at caregivers to relieve some of their burden and support them in the coordination and management of care tasks [6,7].

AAL builds on the classic principles of ambient intelligence (embedded, context-aware, personalized, adaptive, and anticipatory) [8] to create supportive environments for older adults and their caregivers. AAL is an umbrella term for a range of state-of-the-art technologies such as smart home technology, mobile and wearable technology, and assistive robotics [9]. We previously defined AAL as follows [7]:

> *State-of-the-art ICT-based solutions that build on the principles of ambient intelligence to create intelligent environments that provide all-encompassing, non-invasive, and pro-active support to older adults and have the ultimate goal to maintain their independence, enhance their overall quality of life, and support their caregivers.*

Application areas are broad and include, for example, health monitoring, activity monitoring, medication management, fall detection, reminder and planning systems, interactive games and storytelling, care management, social companion robots, and ambient awareness systems.

Although there are high hopes for AAL technologies to solve the challenges of the aging population, different systematic reviews conclude that the technology readiness level of these applications is still low and that most applications have not yet matured into the implementation phase. In addition, scientific evidence for the effectiveness of these technologies is weak and efficiency outcomes are almost nonexistent [6,10,11]. Furthermore, research in the AAL area is still predominately technology oriented [6,12], and there is little theoretical understanding of the user's perspective [11,13]. Hence, there is a need for further research on user acceptance.

The Importance of User Acceptance

User acceptance is key to the successful adoption and diffusion of new technologies. Indeed, several researchers have concluded that understanding user acceptance and incorporating user needs is essential to the successful digitization of the health care sector [14-18]. In the context of AAL, the slow deployment of AAL systems has been attributed to the lack of user acceptance and missing focus on user needs [9,13,19]. Loss of privacy [20-22] and the fear of substituting face-to-face interaction [23-26] are examples of acceptance barriers found in previous research. This is not surprising considering the pervasiveness of these technologies [27]. These applications are designed to be placed in personal environments or directly on the body, collect and store sensitive data, influence behavior and habits, and take over tasks that are usually carried out by the older adults themselves or a human caregiver.

The insufficient understanding of users' needs is also reflected in ageist stereotypes, which are still common in this field [28-30]. These studies portray older adults as a homogeneous group that is frail and lonely and has low technology literacy. To combat these stereotypes, researchers need to adopt a more user-centered mindset and develop a deeper understanding of the user's point of view.

Although the number of studies on user acceptance and user needs has slowly increased over the last couple of years, most research still lacks a solid theoretical foundation to explain and underpin their results [11]. This is also confirmed by Blackman et al [31], who concluded that AAL research is rich in data but poor in theory. A solid theoretical foundation is crucial for understanding the underlying social, psychological, and behavioral mechanisms of the acceptance process. A related concern is the lack of large-scale quantitative research on user acceptance in this area [11,13]. More quantitative approaches are needed to understand the relative importance of acceptance factors, identify their underlying relationships, and make statistically grounded and externally valid inferences about their influence on the acceptance process. Developing a stronger theoretical and statistically grounded understanding of user acceptance in AAL research will improve AAL conceptualization and development. At the same time, it will increase the likelihood of future acceptance by intended users.

The Conceptual Model of AAL Acceptance

Technology acceptance occurs over time and consists of different stages [32-36]. Owing to the overall low maturity of AAL technologies, it was decided to focus on early user acceptance, meaning the factors that contribute to the initial intention to use or reject AAL technology in the future.

Over the years, several theories and models have been developed to explain technology acceptance, including the technology acceptance model (TAM) [37,38], the unified theory of acceptance and use of technology (UTAUT) [39], and the theory of planned behavior (TPB) [40]. Although TAM and UTAUT are popular choices in the field of eHealth [18], we chose TPB as a theoretical foundation for several reasons. First, TPB is a

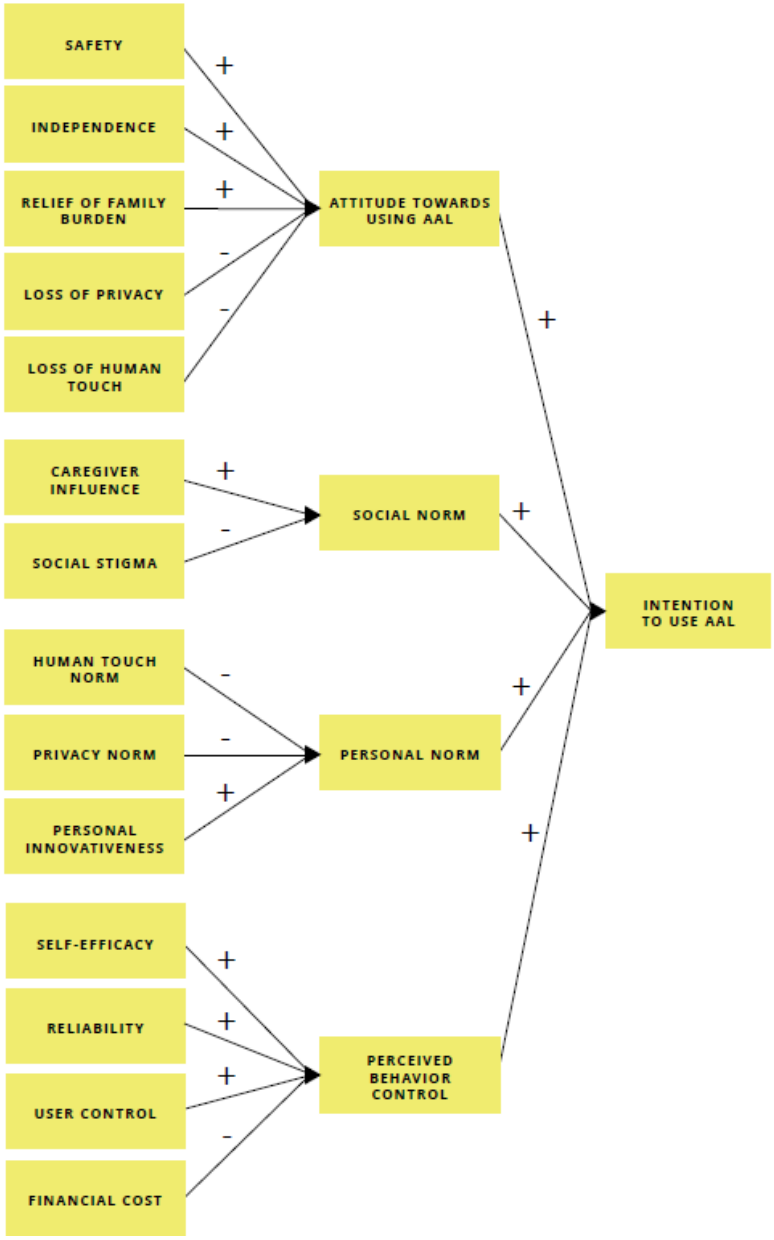
well-known and validated psychological theory to understand and explain human behavior, including technology acceptance [41-44] and health-related behaviors [45,46]. It has also been applied to understand the adoption of assistive devices [47] and eHealth applications [48]. In contrast, UTAUT is an eclectic model that lacks a strong theoretical foundation [49]. Second, TPB provides an ideal basis for understanding early user acceptance by specifically focusing on the attitudinal, social, and normative belief structure that leads to the intention to use a technology. These insights are very informative for further development and implementation of AAL. In contrast, TAM's predominant focus on usefulness and ease of use provides little valuable insights for the design and implementation of new technologies [50]. Third, TPB is explicitly open to the inclusion of more variables [40] and therefore forms a good starting point for developing a new model of AAL acceptance.

Intention is a central construct in TPB and viewed as an immediate determinant of actual behavior. Intention is defined as an "indication of a person's readiness to perform a given behavior." According to TPB, intention is determined by 3 variables: attitude toward the behavior, subjective norm, and perceived behavioral control. Attitude is defined as "the degree to which performance of the behavior is positively or negatively valued." Subjective norm is defined as the "perceived social pressure to engage or not to engage in a behavior." Perceived behavioral control can be described as "people's perceptions of their ability to perform a given behavior." Following an expectancy value approach, in TPB, attitude is determined by a set of behavioral beliefs about the outcome of a given behavior, weighted by the evaluation of that outcome. Subjective norm is determined by a set of normative beliefs concerning the expectations of important referents, weighted by the motivation to comply. Finally, perceived behavioral control is determined by several control beliefs, weighted by its perceived power [40,51,52].

In line with TPB, intention to use *AAL* is proposed as the key dependent variable in our conceptual model, with *attitude toward using AAL*, *social norm*, *and perceived behavioral* control as direct ascendants (Figure 1). Personal norm was added as an additional predictor of intention, thereby answering to the appeal of previous researchers to consider different normative mechanisms for TPB [45,53]. We define personal norm as "people's self-based standards or expectations for AAL use that flow from one's internalized values," thereby referring to Schwartz [54]. The construct was operationalized in terms of self-identity, drawing on the work of Lee et al [55] and Sparks and Shepherd [56].

For the conceptual model, the underlying behavioral, normative, and control belief structures were decomposed into specific multidimensional belief constructs, as suggested by Taylor and Todd [43]. The advantage of this approach is that it emphasizes the relevant beliefs antecedents for AAL acceptance and, consequently, provides more directive insights for the design of AAL technologies [43]. We drew on earlier user research in the field [23,57-59] and our own qualitative user studies [60,61] to select the relevant underlying beliefs. This *resulted in safety*, *independent living*, and *relief of family burden* as positive belief antecedents for attitude and *loss of privacy* and *loss of human touch* as negative belief antecedents. *Caregiver influence* was proposed as a positive antecedent of social norm, whereas social stigma was proposed as a negative antecedent. Personal norm was hypothesized to be positively influenced by one's *personal innovativeness*, whereas human touch norm and privacy norm were suggested as negative antecedents. Finally, perceived behavioral control was hypothesized to be positively influenced by *self-efficacy*, *reliability*, and level of user control and negatively influenced by financial cost. Multimedia Appendix 1 displays an overview of the underlying belief constructs and their definitions.

Figure 1. Conceptual model of ambient assisted living acceptance.



Methods

Overview

A web-based survey of older adults was conducted in the Netherlands to validate the conceptual model. A Dutch ISO (International Organization for Standardization)–certified research agency was hired to distribute the survey. The agency is an expert in web-based fieldwork and manages a panel of 110,000 members with diverse sociodemographic backgrounds.

Participants

Older adults aged between 55 and 85 years were specified as the target population to include older adults with different living and work situations, different perceptions of health and quality of life, different support needs, and different levels of technology experience. Our aim is to adequately represent this highly heterogeneous target group [62,63]. The lower boundary of the age requirement was set at 55 years to include the perspective of the future generation of older adults. Predefined age quotas were used to obtain a representative sample.

Upon invitation, 2113 older adults participated in the survey. Of these participants, 679 participants did not complete the survey, most of whom stopped immediately after the introduction page. Another 138 participants were removed from the sample because of: incomplete response patterns, exceptionally short response times, straight lining, and insufficient understanding of the presented AAL material. This led to a response rate of 61.33% (1296/2113) and a total of 1296 cases for further analysis. The final sample was representative of the older Dutch adult population in terms of age (55-64 years: n=555, 42.82%; 65-74 years: n=497, 38.35%; 75-85 years: n=244, 18.83%) and gender (male: n=637, 49.15%; female: n=659, 50.85%) [64]. Most of the participants (n=1227, 94.68%) had no user experience with AAL applications. All other sample characteristics are presented in Table 1.

Table 1. Sample characteristics (N=1296).

Variables	Values, n (%)
Living situation	
Alone	384 (29.63)
With (partner or family or friend)	912 (70.37)
Education	
Low	474 (36.57)
Intermediate	439 (33.87)
Tertiary	383 (29.55)
Work situation	
Working	332 (25.61)
Not working	960 (74.07)
Self-rated health	
Excellent	88 (6.79)
Very good	245 (18.90)
Good	548 (42.28)
Fair	345 (26.62)
Poor	70 (5.40)
Self-rated quality of life	
Excellent	115 (8.87)
Very good	335 (25.85)
Good	574 (44.29)
Fair	248 (19.14)
Poor	24 (1.85)
Current support need[a]	
No support	1073 (82.79)
Domestic tasks	166 (12.81)
Psychosocial support	88 (6.79)
Personal care	45 (3.47)
Medical care	30 (2.31)
Support provider[a,b]	
Partner	95 (42.60)
Child	62 (27.80)
Family	13 (5.83)
Friend	21 (9.42)
Neighbor	13 (5.83)
Professional	117 (52.47)
Expected support need	
Highly unlikely	150 (11.57)
Less likely than others	186 (14.35)
Equally likely than others	577 (44.52)
More likely than for others	114 (8.80)
Highly likely	90 (6.94)

Variables	Values, n (%)
Don’t know	179 (13.81)

[a]Multiple answers were allowed.

[b]Of those who reported to receive support (n=223).

Survey Materials and Procedure

Participants were presented with a short (2.25 minutes) video animation that explained the concept of AAL [65]. Previous research has shown that animated content with spoken text works well to communicate complex health-related information [66]. For this video animation, a scenario was narrated with the persona Ben, an older adult, and his daughter and informal caregiver, Sophie. Personas and user scenarios are tools that are frequently used in participatory design activities to translate abstract ideas about the user into something more tangible [67]. Three example applications were included in the scenario: (1) smart home technology for activity monitoring and fall detection, (2) a reminder system for appointments and medications, and (3) a social service robot and a social companion robot. In addition to the video animation, participants viewed photos of market-ready AAL products: (1) Sensara activity monitoring [68], (2) Dayclocks reminder application [69], and (3) Zora, a social companion robot [70]. The photos contained a short description of the main features of the product. Two control questions were included to test the understanding of the presented material (“The video/pictures about AAL technology was/were clear to me”). Participants were also asked about their previous knowledge and experience with AAL technology.

After exposure to the video and photos, the participants were directed to the remaining items of the AAL acceptance survey. The survey concluded with questions about the sociodemographic background and participants’ self-rated subjective health and overall quality of life, received level of care, and anticipated need for care in the future.

Measurements

Although some measurements were derived from validated scales, because of the lack of quantitative research in the field, a large part of the measurement was newly developed following the procedure described by DeVellis [71]. Topics from AAL literature and our qualitative user studies [60,61] were used as a starting point to create the initial pool of items. To test and improve the psychometric properties of the newly developed measurements and the overall survey structure, several pretests were conducted. First, the initial pool of items was evaluated for content validity, clarity, and redundancy with 4 senior researchers with expertise in AAL and psychometrics. After this first pretest, some items were removed and the others were rephrased. In the second pretest, the complete web-based survey instrument was presented to 3 older adults to evaluate the overall format (layout, structure, and length), test their ability to navigate through the web-based environment, and evaluate their comprehension of the survey items. Following the guidelines described by Willis [72], we conducted cognitive interviews using a combination of think-aloud and verbal probing techniques, while participants clicked through the survey. As a result, several problem areas were identified and the survey was adjusted accordingly.

We used a 5-point Likert scale as a response scale (1=strongly disagree; 5=strongly agree). For the attitude items, a 5-point semantic differential scale was used. *Don’t know* was included as a response option as AAL is a fairly new concept, and we suspected that some participants would not have a strong enough tendency to formulate an opinion [73]. The *don’t know* option was treated as missing values. Full information maximum likelihood (FIML) was used to deal with the missing values. FIML is considered a robust and state-of-the-art approach to handle missing data and is widely recommended in the methodological literature [74-77]. Table 2 gives a concise overview of the operationalization of the key variables included in the survey instrument. Multimedia Appendix 2 [43,44,47,55,56,78-90] shows the final list of items after validation.

Table 2. Measurements.

Variable name	Number of items in the survey	Example item
Intention to use AAL[a]	4	In the future, I intend to use AAL technology
Attitude toward using AAL	6	I (like/dislike) the idea of using AAL technology
Social norm	3	Most people whose opinion I value, would think positively about my use of AAL technology
Personal norm	3	I view myself as a user of technology for my health and well-being
Perceived behavioral control	4	Using AAL technology is entirely in my control
Safety	6	If I use AAL technology, I will feel safer in my home
Independence	4	If I use AAL technology, I can do things independently
Relief of family burden	6	My use of AAL technology will give my family members peace of mind
Loss of privacy	6	If I use AAL technology, I worry that my personal information might be shared with others without my permission
Loss of human touch	6	If I use AAL technology, I will get less personal attention
Caregiver influence	3	My caregivers would have a positive view on my use of AAL technology
Social stigma	4	If I use AAL technology, I am concerned that the technology will be visible to others
Human touch norm	4	I prefer personal care over care via AAL technology
Privacy norm	6	I think I have the right to control my personal information
Personal innovativeness	4	If I heard about a new information technology, I would look for ways to experiment with it
Self-efficacy	7	If I had problems relating to using AAL technology I know I could work them out
User control	3	I think that I will feel in control, when using AAL technology
Reliability	4	I think that AAL technology is reliable
Financial cost	3	I think that using AAL technology will be expensive

[a]AAL: ambient assisted living.

Structural Equation Modeling

We used structural equation modeling (SEM) to validate the conceptual model.

The measurement model was validated in 2 stages. First, a pilot study was conducted among 320 older adults in the Netherlands. The hypothesized relationships between the latent variables and their indicator variables were explored using confirmatory factor analysis. Although this technique is labeled as confirmatory, it was used in an exploratory and iterative manner by paying attention to the posthoc modification indices [91]. By specifying the relationships between the latent variables and their indicator variables a priori, we employed a theory-driven approach rather than a data-driven approach to validate the measurements [92,93]. The measurement model was respecified with the main study sample (N=1296), leading to further refinement of the measurement model.

We used the Lavaan package version 0.5-23 [94] in R version 3.4.3 [95] to perform the analysis. Maximum likelihood estimation with FIML for missing data was used because the data were approximately normally distributed. The original measurement model proposed 19 distinct latent factors and 86 indicator variables. Indicators with poor standardized factor loading (<0.50) and low squared multiple correlation (SMC<0.40) were removed. To further evaluate the convergent validity of the measurement model, we assessed the McDonald hierarchical omega [96], Cronbach alpha [97], and the average variance extracted (AVE) for each latent variable. The threshold for the former 2 measurements was 0.70, and the recommended AVE value threshold was 0.50. Discriminant validity was examined using the heterotrait-monotrait (HTMT) ratio. If the HTMT value is <0.90, discriminant validity is established [98]. After validating the measurement model, the structural equation model was tested.

Results

Measurement Model

The fit measures of the original model were less than aspiration values. After the inspection of factor loadings and SMC values, several indicators were iteratively removed. This included the latent variable user control, as 2 of the 3 indicators loaded poorly on the latent construct. A minimum of 3 indicators are required to represent the latent variable [99]. The indicators of the latent variable privacy norm had low or just acceptable SMC values. As the variable showed relatively weak psychometric properties across the 2 independent samples, privacy norm was removed from the measurement model. One indicator (PSN03) with an

SMC value less than the aspiration value was not excluded to meet the requirement of the 3 indicators to represent the latent variable. Another indicator less than the aspiration value (PI02) was included because it originated in a validated scale [78]. Upon inspection of the posthoc modification indexes, suggested residual correlations between the following indicator pairs were added: PSN03 and PI02, ATT02 and ATT03, ATT04 and ATT05, LP03 and LP05, LP03 and LP06, LP05 and LP06, and FB03 and FB05. After calculating the hierarchical omega, Cronbach alpha, and AVE values, it was decided to remove the latent variable social stigma from the measurement model because of a low AVE value (AVE=0.47) and overall weak psychometric properties across the 2 samples. Finally, HTMT values indicated that safety and independence should be considered as a single latent variable called *safe and independent living.*

The final measurement model consisted of 15 latent factors, 63 indicators, and 7 added residual correlations. The model showed acceptable-to-good fit for all fit measures (root mean square error of approximation [RMSEA]=0.04; standardized root mean square residual [SRMR]=0.05; comparative fit index [CFI]=0.93; and Tucker Lewis index [TLI]=0.92). Multimedia Appendix 3 displays the final list of indicators with intercept (FIML mean), indicator mean (values with listwise deletion), SD (values with listwise deletion), factor loadings, SMC, hierarchical omega, Cronbach alpha, and AVE.

Descriptives

The indicator scores from the final measurement model were pooled into a composite score for each latent variable. Table 3 shows an overview of the composite mean, SD, and range for each latent variable.

Table 3. Composite mean and SD per latent variable.[a]

Latent variable	Mean (SD)	Minimum	Maximum
Intention to use AAL[b]	3.34 (0.73)	1	5
Attitude toward using AAL	3.73 (0.78)	1	5
Social norm	3.67 (0.57)	1	5
Personal norm	3.42 (0.75)	1	5
Perceived behavioral control	3.32 (0.71)	1	5
Safe and independent living	3.92 (0.52)	1	5
Relief of family burden	3.67 (0.65)	1	5
Loss of privacy	3.14 (0.87)	1	5
Loss of human touch	3.13 (0.83)	1	5
Caregiver influence	3.73 (0.56)	1	5
Human touch norm	3.97 (0.67)	1	5
Personal innovativeness	3.19 (0.78)	1	5
Self-efficacy	3.79 (0.60)	1	5
Reliability	3.26 (0.59)	1	5
Financial cost	3.81 (0.68)	1	5

[a]Single imputation with the Expectation Maximization method was used to handle the missing data for the composite scores and group comparison.

[b]AAL: ambient assisted living.

The overall intention to use AAL technology was moderately high in the sample (mean 3.34, SD 0.73). This means that, in general, older adults were relatively open to the idea of using AAL technologies in the future. Regarding the 3 age quotas, there was no significant difference in their use intention ($F_{2,1293}$=2.89; *P*=.06). Similarly, we found no significant differences across different levels of subjective health ($F_{4,1291}$=0.60; *P*=.66) and expected support needs ($F_{4,1112}$=0.52; *P*=.72).

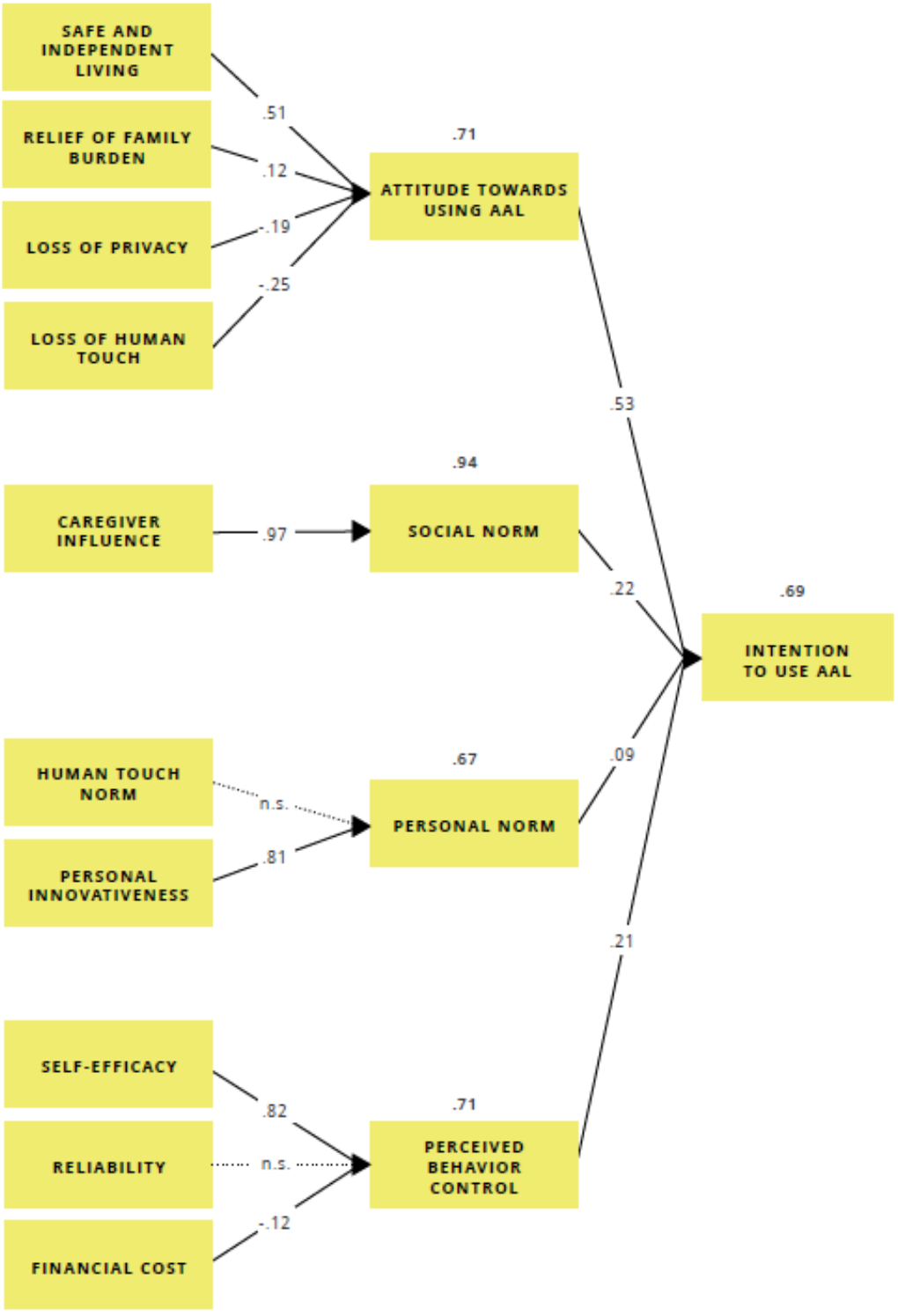
Structural Equation Model

The hypothesized structural equation model showed good overall fit with the observed data: RMSEA=0.04, SRMR=0.06, CFI=0.93, and TLI=0.92. The model accounted for 69% of the variance in the intention to use AAL (R^2=0.69). All but 2 of the hypothesized paths had significant standardized path coefficients at the *P*<.001 level.

Attitude toward using AAL, social norm, personal norm, and perceived behavior control significantly affected the intention to use AAL. Attitude was the most important influencer of intention (β=.53). Attitude toward using AAL was affected by older adults' expectations about safe and independent living (β=.51), relief of family burden (β=.12), loss of privacy (β=–.19), and loss of human touch (β=−.25). Together, these variables explained 71% of the variance in attitude (R^2=0.71). Social norm was strongly affected by caregiver influence (β=.97). Caregiver influence predicted 94% of the variance in social norm. The hypothesized influence of human touch norm on personal norm was not significant (*P*=.39), and personal innovativeness therefore remained to be the only significant

predictor of personal norm (β=.81). Personal innovativeness explained 67% of the variance in personal norm. Self-efficacy (β=.81) and financial cost (β=−.12) remained to be the 2 predictors of perceived behavior control. Together, these variables explained 71% of the variance in the perceived behavior control. The expected influence of reliability was not significant (*P*=.68; Figure 2).

Figure 2. Structural equation model. Values adjacent to the single-headed arrows represent the standardized regression coefficients (*P*<.001). The dotted lines represent the nonsignificant paths. Values above the variable rectangles represent the variance explained in the latent variables.



Discussion

Principal Findings

The aim of this study is to develop a statistically grounded understanding of the acceptance of AAL technology among older adults in the Netherlands. Specifically, this study aimed to compare the relative importance of different acceptance factors, their underlying relationships, and their explanatory power for the intention to use AAL technologies in the future.

The results of the web-based acceptance survey showed that the proposed model of AAL acceptance showed a good model fit for the observed data and explained 69% of the variance in intention to use. All hypothesized paths were significant, except for the path between human touch norm and personal norm and the path between reliability and perceived behavior control. Therefore, it can be concluded that our established theoretical model provides a valuable framework for understanding and explaining older adults' acceptance in the early acceptance stage.

The overall intention to use AAL technology was moderately positive. This means that older adults are relatively open to the idea of using AAL technologies in the future. We found no difference in intention to use between age groups, people with different subjective health ratings, and different expected support needs. Although this might be somewhat surprising, this is in line with findings from Ziefle and Röcker [100], who found that age and subjective health status did not influence the willingness to use AAL technologies.

As expected, the intention to use AAL was predicted by attitude toward using AAL, social norm, personal norm, and perceived behavior control. Attitude was the most important predictor, followed by social norm and perceived behavior control. The results showed only a weak influence on personal norm. Ajzen [40] argues that "the relative importance of attitude, subjective norm, and perceived behavioral control in the prediction of intention is expected to vary across behaviors and situations." From the results, we can conclude that in an early acceptance stage, in which people have no or limited experience with AAL technologies, the overall attitude toward using AAL is the most important influencer of intention to use. On the other hand, self-based standards and expectations regarding AAL use are only minor influencers of older adults' intention to use.

Safe and independent living was the most important positive influencer of attitude, which in turn influenced intention to use. This is in line with previous AAL research [57,59,101] and our own qualitative user studies [60,61]. Older adults regarded the increased feeling of safety and the opportunity of independent living as a major advantage of AAL. We also found empirical evidence that promises of safety and autonomy are a valid trade-off for concerns about personal interaction and privacy, as suggested by earlier research [58,102]. Nevertheless, in line with previous studies [20,21,23-26], both concerns still substantially contributed to a negative attitude toward using AAL and should be considered when developing AAL applications. Earlier studies [58,101,103] suggested that older adults perceive AAL technologies as good tools for reducing the overall burden on caregivers. This was confirmed by the results of the AAL acceptance survey.

Previous research has suggested that the influence of caregivers, especially informal family caregivers, is important for the acceptance of AAL technologies [104-106]. Although we did not distinguish between formal and informal caregiver influence, the findings of the AAL acceptance survey indeed identified caregivers as crucial social referents for building social norm. Social norm, in turn, influenced use intentions. For future research, it would be interesting to explicitly distinguish between formal and informal caregiver influence.

In line with our qualitative user studies [60,61], older adults' general willingness to try out new information technology positively contributed to their overall personal norm. However, the effect of personal norm on intention to use AAL was weak in the current sample. In contrast to our expectations, human touch norm had no significant influence on personal norm. An explanation for this finding may be that older adults preferred human care over care via AAL technology but could still identify as users of AAL.

Self-efficacy is a concept derived from social cognitive theory and is an essential determinant of human motivation and behavior [107,108]. Following previous research [57], it was hypothesized that self-efficacy would positively affect use intention via perceived behavior control. This hypothesized relationship was confirmed through the results of the AAL acceptance survey. Moreover, in line with previous research [57,101], expectations about high financial cost negatively contributed to perceived behavioral control. The hypothesized relationship between perceived behavior control and reliability was not significant. We suspect that with no or limited experience of AAL, users found it difficult to formulate specific and consistent expectations about the expected reliability of AAL. However, we believe that reliability will be considered in a later acceptance stage when users are actively interacting with the technology. Therefore, future research should consider these variables.

Limitations

As in every study, there were some limitations to be considered. First, by using a web-based survey, we accepted that our sample had a bias toward older adults with internet connection and some technology skills. However, most older adults are active internet users [109,110]. Hence, the current sample remains to be largely representative of the older Dutch adult population. Second, participants' responses were based on the provided study material and not on direct interaction with AAL technologies. This could have limited participants' impressions of AAL. However, this fits the phase of early acceptance. In real-life situations, older adults will not necessarily try out a new technological device before forming their initial use intention. Previous research has shown that participants can form attitudes and expectations toward new and unfamiliar technologies without active use experience [32,111-113]. Finally, reaching an acceptable model fit in SEM does not imply that the hypothesized model is the only fitting model. Other equivalent or near-equivalent models may show equal or even better fit [93]. However, at this stage, the AAL field does not

offer a rich theoretical discourse to inspire alternative models. Moreover, the measurement part of the model was cross-validated across 2 independent samples. In addition, the model was built on a strong and well-established psychological theory (TPB), a literature review, and several qualitative user studies.

Future Research

To the best of our knowledge, this model is one of the first theory-driven quantitative frameworks for understanding AAL acceptance, which has been validated with a representative sample of the target population. However, this also means that this model is the first approximation to explain AAL acceptance. Further cross-validation and refinement is needed to ensure that this model remains stable and valid across different populations and cultural contexts. The established model focuses on early user acceptance and the initial intention to use AAL. Future research needs to implement longitudinal designs to explore later stages of acceptance when older adults start using the technology in their own home environment and attitudes, user needs, and intentions might change [34,35,114]. This study focused on older adults. Other important stakeholder groups include informal and formal caregivers. They can be primary users of AAL applications [115] and are important in signaling the older adults' need for support and introducing AAL into the home care practice [7]. Hence, future research should further investigate caregivers' perceptions of AAL.

For now, our insights into early acceptance among older adults can shape the further discourse and implementation of AAL.

Conclusions

For the future success of AAL, it is vital to know if these technologies will fall on fertile ground and will be accepted by the intended users. In other words, will the policy vision of AAL as a solution to healthy and independent aging become reality from the perspective of older adult users? This study shows that Dutch older adults seem receptive to the idea of using AAL technology in the future. Being mindful of the acceptance factors will help developers make more informed design decisions before diffusing applications into the market.

Although the provided model focuses on AAL technologies, our insights on acceptance factors (eg, loss of privacy, loss of human touch, caregiver influence, financial cost) can also be valuable for the broader field of eHealth development and implementation.

Acknowledgments

The authors would like to thank the participants for taking the time to participate in the survey.

Authors' Contributions

CJ, SBA, JD, and OP were responsible for the conception and design of the study. CJ managed the data collection. All authors were involved in the statistical analysis and interpretation of the data. CJ, SBA, and OP drafted and critically revised the manuscript. CJ and SBA approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definition of the proposed belief constructs.
[PDF File (Adobe PDF File), 401 KB - jmir_v23i3e22613_app1.pdf]

Multimedia Appendix 2
Refined measurements of the ambient assisted living acceptance survey.
[PDF File (Adobe PDF File), 573 KB - jmir_v23i3e22613_app2.pdf]

Multimedia Appendix 3
Final measurement model intercept, mean, SD, standardized factor loadings, hierarchical omega, Cronbach alpha, and average variance extracted.
[PDF File (Adobe PDF File), 572 KB - jmir_v23i3e22613_app3.pdf]

References

1. Hasson H, Arnetz JE. Nursing staff competence, work strain, stress and satisfaction in elderly care: a comparison of home-based care and nursing homes. J Clin Nurs 2008 Feb;17(4):468-481. [doi: 10.1111/j.1365-2702.2006.01803.x] [Medline: 17331093]
2. Deeken JF, Taylor KL, Mangan P, Yabroff KR, Ingham JM. Care for the caregivers: a review of self-report instruments developed to measure the burden, needs, and quality of life of informal caregivers. J Pain Symptom Manage 2003 Oct;26(4):922-953 [FREE Full text] [doi: 10.1016/s0885-3924(03)00327-0] [Medline: 14527761]

3. Colombo F, Llena-Nozal A, Mercier J, Tjadens F. Help Wanted? Providing and Paying for Long-Term Care. OECD Publishing. 2011 May 18. URL: https://www.oecd-ilibrary.org/social-issues-migration-health/help-wanted_9789264097759-en [accessed 2020-07-17]
4. Active Ageing: a Policy Framework. World Health Organization. 2002 Apr. URL: https://www.who.int/ageing/publications/active_ageing/en/ [accessed 2020-07-17]
5. Foster L, Walker A. Active and successful aging: a European policy perspective. Gerontologist 2015 Feb;55(1):83-90 [FREE Full text] [doi: 10.1093/geront/gnu028] [Medline: 24846882]
6. Calvaresi D, Cesarini D, Sernani P, Marinoni M, Dragoni AF, Sturm A. Exploring the ambient assisted living domain: a systematic review. J Ambient Intell Human Comput 2016 May 4;8(2):239-257. [doi: 10.1007/s12652-016-0374-3]
7. Jaschinski C, Ben Allouch S. Listening to the ones who care: exploring the perceptions of informal caregivers towards ambient assisted living applications. J Ambient Intell Human Comput 2018 May 25;10(2):761-778. [doi: 10.1007/s12652-018-0856-6]
8. Aarts EHL, Marzano S. The New Everyday: Views on Ambient Intelligence. Rotterdam: 010 Publishers; 2003.
9. Rashidi P, Mihailidis A. A survey on ambient-assisted living tools for older adults. IEEE J Biomed Health Inform 2013 May;17(3):579-590. [doi: 10.1109/jbhi.2012.2234129] [Medline: 24592460]
10. Krick T, Huter K, Domhoff D, Schmidt A, Rothgang H, Wolf-Ostermann K. Digital technology and nursing care: a scoping review on acceptance, effectiveness and efficiency studies of informal and formal care technologies. BMC Health Serv Res 2019 Jun 20;19(1):400 [FREE Full text] [doi: 10.1186/s12913-019-4238-3] [Medline: 31221133]
11. Liu L, Stroulia E, Nikolaidis I, Miguel-Cruz A, Rios Rincon A. Smart homes and home health monitoring technologies for older adults: A systematic review. Int J Med Inform 2016 Jul;91:44-59. [doi: 10.1016/j.ijmedinf.2016.04.007] [Medline: 27185508]
12. Queirós A, Silva A, Alvarelhão J, Rocha NP, Teixeira A. Usability, accessibility and ambient-assisted living: a systematic literature review. Univ Access Inf Soc 2013 Oct 5;14(1):57-66. [doi: 10.1007/s10209-013-0328-x]
13. Peek STM, Wouters EJM, van HJ, Luijkx KG, Boeije HR, Vrijhoef HJ. Factors influencing acceptance of technology for aging in place: a systematic review. Int J Med Inform 2014 Apr;83(4):235-248 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.01.004] [Medline: 24529817]
14. McGinn CA, Grenier S, Duplantie J, Shaw N, Sicotte C, Mathieu L, et al. Comparison of user groups' perspectives of barriers and facilitators to implementing electronic health records: a systematic review. BMC Med 2011 Apr 28;9:46 [FREE Full text] [doi: 10.1186/1741-7015-9-46] [Medline: 21524315]
15. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful optimization of large-scale health information technology. J Am Med Inform Assoc 2016 Apr 23. [doi: 10.1093/jamia/ocw037] [Medline: 27107441]
16. Cresswell K, Morrison Z, Crowe S, Robertson A, Sheikh A. Anything but engaged: user involvement in the context of a national electronic health record implementation. Inform Prim Care 2011;19(4):191-206 [FREE Full text] [doi: 10.14236/jhi.v19i4.814] [Medline: 22828574]
17. Yusif S, Hafeez-Baig A, Soar J. e-Health readiness assessment factors and measuring tools: a systematic review. Int J Med Inform 2017 Dec;107:56-64. [doi: 10.1016/j.ijmedinf.2017.08.006] [Medline: 29029692]
18. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
19. Robinson H, MacDonald B, Broadbent E. The Role of Healthcare Robots for Older People at Home: A Review. Int J of Soc Robotics 2014 Jul 3;6(4):575-591. [doi: 10.1007/s12369-014-0242-2]
20. Yusif S, Soar J, Hafeez-Baig A. Older people, assistive technologies, and the barriers to adoption: a systematic review. Int J Med Inform 2016 Dec;94:112-116. [doi: 10.1016/j.ijmedinf.2016.07.004] [Medline: 27573318]
21. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014 Sep;83(9):624-635 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
22. Wilkowska W, Ziefle M. Privacy and data security in E-health: requirements from the user's perspective. Health Informatics J 2012 Sep;18(3):191-201 [FREE Full text] [doi: 10.1177/1460458212442933] [Medline: 23011814]
23. Beer JM, Takayama L. Mobile Remote Presence Systems for Older Adults: Acceptance, Benefits, and Concerns. In: Proceedings of the 6th international conference on Human-robot interaction. New York: ACM; 2011 Presented at: 6th International Conference on Human-Robot Interaction (HRI '11); 2011 March; Lausanne, Switzerland p. 19-26. [doi: 10.1145/1957656.1957665]
24. Demiris G, Rantz M, Aud M, Marek K, Tyrer H, Skubic M, et al. Older adults' attitudes towards and perceptions of 'smart home' technologies: a pilot study. Med Inform Internet Med 2004 Jun;29(2):87-94. [doi: 10.1080/14639230410001684387] [Medline: 15370989]
25. Marquis-Faulkes F, McKenna SJ, Gregor P, Newell AF. Scenario-based drama as a tool for investigating user requirements with application to home monitoring for elderly people. In: Proceedings of the 10th International Conference on Human-Computer Interaction. 2003 Presented at: 10th International Conference on Human-Computer Interaction (HCI International); 2003 June 22-27; Crete, Greece p. 512-516.

26. Kang HG, Mahoney DF, Hoenig H, Hirth VA, Bonato P, Hajjar I, et al. In situ monitoring of health in older adults: technologies and issues. J Am Geriatr Soc 2010 Aug;58(8):1579-1586. [doi: 10.1111/j.1532-5415.2010.02959.x] [Medline: 20646105]
27. Novitzky P, Smeaton AF, Chen C, Irving K, Jacquemard T, O'Brolcháin F, et al. A review of contemporary work on the ethics of ambient assisted living technologies for people with dementia. Sci Eng Ethics 2015 Jun;21(3):707-765. [doi: 10.1007/s11948-014-9552-x] [Medline: 24942810]
28. Fischer B, Peine A, Östlund B. The Importance of User Involvement: A Systematic Review of Involving Older Users in Technology Design. Gerontologist 2020 Sep 15;60(7):e513-e523 [FREE Full text] [doi: 10.1093/geront/gnz163] [Medline: 31773145]
29. Peine A, Rollwagen I, Neven L. The rise of the 'innosumer'—Rethinking older technology users. Technological Forecasting and Social Change 2014 Feb;82:199-214. [doi: 10.1016/j.techfore.2013.06.013]
30. Vines J, Pritchard G, Wright P, Olivier P, Brittain K. An Age-Old Problem: Examining the Discourses of Ageing in HCI and Strategies for Future Research. ACM Trans. Comput.-Hum. Interact 2015 Mar 04;22(1):1-27. [doi: 10.1145/2696867]
31. Blackman S, Matlo C, Bobrovitskiy C, Waldoch A, Fang M, Jackson P, et al. Ambient Assisted Living Technologies for Aging Well: A Scoping Review. Journal of Intelligent Systems 2016;25(1):55-69. [doi: 10.1515/jisys-2014-0136]
32. Karahanna E, Straub DW, Chervany NL. Information Technology Adoption Across Time: A Cross-Sectional Comparison of Pre-Adoption and Post-Adoption Beliefs. MIS Q 1999;23(2):183-213. [doi: 10.2307/249751]
33. Karapanos E, Zimmerman J, Forlizzi J, Martens J. User experience over time: An initial framework. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: ACM; 2009 Presented at: SIGCHI Conference on Human Factors in Computing Systems (CHI '09); 2009 April; Boston, MA, USA p. 729-738. [doi: 10.1145/1518701.1518814]
34. de Graaf MM, Ben Allouch S, van Dijk JAGM. Long-term evaluation of a social robot in real homes. Interaction Studies 2016 Jan;17(3):462-491. [doi: 10.1075/is.17.3.08deg]
35. Chiu TML, Eysenbach G. Stages of use: consideration, initiation, utilization, and outcomes of an internet-mediated intervention. BMC Med Inform Decis Mak 2010;10:73 [FREE Full text] [doi: 10.1186/1472-6947-10-73] [Medline: 21092275]
36. Rogers EM. Diffusion of Innovations. 5th edition. New York: Free Press; 2003.
37. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
38. Davis FD. A technology acceptance model for empirically testing new end-user information systems : theory and results. Cambridge: Massachusetts Institute of Technology; 1986. URL: http://hdl.handle.net/1721.1/15192 [accessed 2020-07-17]
39. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]
40. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
41. Pavlou PA, Fygenson M. Understanding and Predicting Electronic Commerce Adoption: An Extension of the Theory of Planned Behavior. MIS Quarterly 2006 Mar;30(1):115-143. [doi: 10.2307/25148720]
42. de Graaf MMA, Ben Allouch S, van Dijk JAGM. Why Would I Use This in My Home? A Model of Domestic Social Robot Acceptance. Human–Computer Interaction 2017 Jul 21;34(2):115-173. [doi: 10.1080/07370024.2017.1312406]
43. Taylor S, Todd PA. Understanding information technology usage: A test of competing models. Inf Syst Res 1995;6(2):144-176. [doi: 10.1287/isre.6.2.144]
44. Mathieson K. Predicting user intentions: Comparing the technology acceptance model with the theory of planned behavior. Inf Syst Res 1991;2(3):173-191. [doi: 10.1287/isre.2.3.173]
45. Godin G, Kok G. The theory of planned behavior: a review of its applications to health-related behaviors. Am J Health Promot 1996;11(2):87-98. [doi: 10.4278/0890-1171-11.2.87] [Medline: 10163601]
46. Hackman CL, Knowlden AP. Theory of reasoned action and theory of planned behavior-based dietary interventions in adolescents and young adults: a systematic review. Adolesc Health Med Ther 2014;5:101-114 [FREE Full text] [doi: 10.2147/AHMT.S56207] [Medline: 24966710]
47. Roelands M, Van Oost P, Depoorter A, Buysse A. A social-cognitive model to predict the use of assistive devices for mobility and self-care in elderly people. Gerontologist 2002 Feb;42(1):39-50. [doi: 10.1093/geront/42.1.39] [Medline: 11815698]
48. Herrmann LK, Kim J. The fitness of apps: a theory-based examination of mobile fitness app usage over 5 months. Mhealth 2017;3:2 [FREE Full text] [doi: 10.21037/mhealth.2017.01.03] [Medline: 28293619]
49. Bagozzi RP. The Legacy of the Technology Acceptance Model and a Proposal for a Paradigm Shift. Journal of the Association for Information Systems 2007 Apr;8(4):244-254. [doi: 10.17705/1jais.00122]
50. Benbasat I, Barki H. Quo vadis TAM? Journal of the Association for Information Systems 2007 Apr;8(4):211-218.
51. Ajzen I, Fishbein M. The Influence of Attitudes on Behavior. In: Albarracin D, Johnson BT, Zanna MP, editors. The handbook of attitudes. New Jersey: Lawrence Erlbaum Associates Publishers; 2005:173-221.

52. Ajzen I. Theory of Planned Behavior Diagram. 2019. URL: https://people.umass.edu/aizen/tpb.diag.html#null-link [accessed 2020-07-03]
53. Armitage CJ, Conner M. Efficacy of the Theory of Planned Behaviour: a meta-analytic review. Br J Soc Psychol 2001 Dec;40(Pt 4):471-499. [Medline: 11795063]
54. Schwartz SH. Normative Influences on Altruism. Advances in Experimental Social Psychology 1977;10:221-279. [doi: 10.1016/s0065-2601(08)60358-5]
55. Lee Y, Lee J, Lee Z. Social Influence on Technology Acceptance Behaviorlf-Identity Theory Perspective). SIGMIS Database 2006 Sep 19;37(2-3):60-75. [doi: 10.1145/1161345.1161355]
56. Sparks P, Shepherd R. Self-Identity and the Theory of Planned Behavior: Assesing the Role of Identification with Green Consumerism. Social Psychology Quarterly 1992 Dec;55(4):388-399. [doi: 10.2307/2786955]
57. Steele R, Lo A, Secombe C, Wong YK. Elderly persons' perception and acceptance of using wireless sensor networks to assist healthcare. Int J Med Inform 2009 Dec;78(12):788-801. [doi: 10.1016/j.ijmedinf.2009.08.001] [Medline: 19717335]
58. Wild K, Boise L, Lundell J, Foucek A. Unobtrusive in-home monitoring of cognitive and physical health: reactions and perceptions of older adults. J Appl Gerontol 2008;27(2):181-200 [FREE Full text] [doi: 10.1177/0733464807311435] [Medline: 19165352]
59. van Hoof J, Kort HSM, Rutten PGS, Duijnstee MSH. Ageing-in-place with the use of ambient intelligence technology: perspectives of older users. Int J Med Inform 2011 May;80(5):310-331 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.02.010] [Medline: 21439898]
60. Jaschinski C, Ben Allouch S. An Extended View on Benefits and Barriers of Ambient Assisted Living Solutions. International Journal on Advances in Life Sciences 2015;7(1&2):40-53.
61. Jaschinski C, Ben Allouch S. Understanding the User's Acceptance of a Sensor-Based Ambient Assisted Living Application. In: Salah AA, Kröse BJA, Cook DJ, editors. Human Behavior Understanding. Lecture Notes in Computer Science, vol 9277. Cham: Springer International Publishing; Nov 2015:13-25. [doi: 10.1007/978-3-319-24195-1_2]
62. Gregor P, Newell AF, Zaijcek M. Designing for Dynamic Diversity: Interfaces for Older People. In: Proceedings of the 5th International ACM Conference on Assistive Technologies. New York: ACM; 2002 Jul Presented at: 5th International ACM Conference on Assistive Technologies (Assets '02); 2002 July; Edinburgh, Scotland p. 151-156. [doi: 10.1145/638249.638277]
63. Grates M, Heming A, Vukoman M, Schabsky P, Sorgalla J. New Perspectives on User Participation in Technology Design Processes: An Interdisciplinary Approach. Gerontologist 2019 Jan 09;59(1):45-57. [doi: 10.1093/geront/gny112] [Medline: 30260407]
64. Statistics Netherlands. StatLine - Bevolking; geslacht, leeftijd en burgerlijke staat, 1 januari. 2020. URL: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/7461bev/table?dl=B875 [accessed 2020-07-17]
65. Jaschinski C. Independent Aging with the Help of Smart Technology. 2018 Apr 26. URL: https://www.youtube.com/watch?v=TZfy5KW9kOY [accessed 2020-07-17]
66. Meppelink CS, van Weert JCM, Haven CJ, Smit EG. The effectiveness of health animations in audiences with different health literacy levels: an experimental study. J Med Internet Res 2015 Jan 13;17(1):e11 [FREE Full text] [doi: 10.2196/jmir.3979] [Medline: 25586711]
67. Grudin J, Pruitt J. Personas, Participatory Design and Product Development: An Infrastructure for Engagement. In: Proceedings of Participation and Design Conference. 2002 Presented at: Participation and Design Conference (PDC 2002); 2002 June 23-25; Malmo, Sweden p. 144-152. [doi: 10.1145/997078.997089]
68. Sensara. Sensara: Beter weten, beter zorgen. URL: https://sensara.eu/#start-0 [accessed 2020-06-20]
69. DayClocks International B.V. DayClocks - Dé tablet voor oriëntatie, planning en verbinding. URL: https://www.dayclocks.nl/ [accessed 2020-06-20]
70. Zora Robotics NV. Zorabots. URL: https://www.zorarobotics.be [accessed 2020-06-20]
71. DeVellis R. Scale development:Theory and Applications. Thousand Oaks: SAGE Publications; 2003.
72. Willis GB. Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks: SAGE Publications; 2005.
73. Rubin DB, Stern HS, Vehovar V. Handling Don't Know Survey Responses: The Case of the Slovenian Plebiscite. Journal of the American Statistical Association 1995 Sep;90(431):822-828. [doi: 10.2307/2291315]
74. Baraldi AN, Enders CK. An introduction to modern missing data analyses. J Sch Psychol 2010 Feb;48(1):5-37. [doi: 10.1016/j.jsp.2009.10.001] [Medline: 20006986]
75. Little TD, Jorgensen TD, Lang KM, Moore EWG. On the joys of missing data. J Pediatr Psychol 2014 Mar 08;39(2):151-162. [doi: 10.1093/jpepsy/jst048] [Medline: 23836191]
76. Enders C, Bandalos D. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. Structural Equation Modeling 2001 Jul 1;8(3):430-457. [doi: 10.1207/s15328007sem0803_5]
77. Olinsky A, Chen S, Harlow L. The comparative efficacy of imputation methods for missing data in structural equation modeling. European Journal of Operational Research 2003 Nov;151(1):53-79. [doi: 10.1016/s0377-2217(02)00578-7]
78. Agarwal R, Prasad J. A Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology. Inf Syst Res 1998;9(2):204-215. [doi: 10.1287/isre.9.2.204]

79. Ben Allouch S, van Dijk JAGM, Peters O. The acceptance of domestic ambient intelligence appliances by prospective users. In: Tokuda H, Beigl M, Friday M, Brush AJB, Tobe Y, editors. Pervasive Computing. Pervasive 2009. Lecture Notes in Computer Science, vol 5538. Berlin, Heidelberg: Springer; 2009. [doi: 10.1007/978-3-642-01516-8_7]
80. Norman P, Bell R. The Theory of Planned Behaviour and exercise: Evidence for the moderating role of past behaviour. Br J Health Psychol 2010 Dec 16;5(3):249-261. [doi: 10.1348/135910700168892]
81. Bédard M, Molloy DW, Squire L, Dubois S, Lever JA, O'Donnell M. The Zarit Burden Interview: a new short version and screening version. Gerontologist 2001 Oct;41(5):652-657. [doi: 10.1093/geront/41.5.652] [Medline: 11574710]
82. Boise L, Wild K, Mattek N, Ruhl M, Dodge HH, Kaye J. Willingness of older adults to share data and privacy concerns after exposure to unobtrusive in-home monitoring. Gerontechnology 2013;11(3):428-435 [FREE Full text] [doi: 10.4017/gt.2013.11.3.001.00] [Medline: 23525351]
83. Kirchbuchner F, Grosse-Puppendahl T, Hastall MR, Distler M, Kuijper A. Ambient intelligence from senior citizens' perspectives: understanding privacy concerns, technology acceptance, and expectations. In: Ambient Intelligence. AmI 2015. Lecture Notes in Computer Science. Cham: Springer; 2015. [doi: 10.1007/978-3-319-26005-1_4]
84. Ajzen I. Constructing a theory of planned behavior questionnaire. TPB Questionnaire Construction. URL: http://people.umass.edu/aizen/pdf/tpb.measurement.pdf [accessed 2021-03-03]
85. Dabholkar P. Consumer evaluations of new technology-based self-service options: an investigation of alternative models of service quality. Int J Res Marketing 1996 Feb;13(1):29-51. [doi: 10.1016/0167-8116(95)00027-5]
86. Phang C, Sutanto J, Kankanhalli A, Li Y, Tan B, Teo H. Senior citizens' acceptance of information systems: a study in the context of e-government services. IEEE Trans Eng Manage 2006 Nov;53(4):555-569. [doi: 10.1109/TEM.2006.883710]
87. Larose R. Measuring sustainable broadband adoption: an innovative approach to understanding broadband adoption and use. Int J Commun 2012 Jan:2576-2600 [FREE Full text] [doi: 10.5167/uzh-71046]
88. Meuter ML, Ostrom AL, Bitner MJ, Roundtree R. The influence of technology anxiety on consumer use and experiences with self-service technologies. J Bus Res 2003;56(11):899-906. [doi: 10.1016/s0148-2963(01)00276-4]
89. Mcknight DH, Carter M, Thatcher JB, Clay PF. Trust in a specific technology. ACM Trans Manage Inf Syst 2011 Jun;2(2):1-25. [doi: 10.1145/1985347.1985353]
90. Luarn P, Lin H. Toward an understanding of the behavioral intention to use mobile banking. Comput Human Behav 2005 Nov;21(6):873-891. [doi: 10.1016/j.chb.2004.03.003]
91. Bentler PM. SEM with simplicity and accuracy. J Consum Psychol 2010 Apr;20(2):215-220 [FREE Full text] [doi: 10.1016/j.jcps.2010.03.002] [Medline: 20454589]
92. Jöreskog KG. Testing Structural Equation Models. In: Bollen KA, Long JS, editors. Testing Structural Equation Models. Newbury Park: SAGE Publications; 1993:294-316.
93. Kline RB. Principles and practice of structural equation modeling. New York: Guilford Press; 2016.
94. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software 2012;48(2):1-36. [doi: 10.18637/jss.v048.i02]
95. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing URL: https://www.r-project.org [accessed 2020-07-02]
96. McDonald RP, Ho MR. Principles and practice in reporting structural equation analyses. Psychol Methods 2002 Mar;7(1):64-82. [Medline: 11928891]
97. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/BF02310555]
98. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 2014 Aug 22;43(1):115-135. [doi: 10.1007/s11747-014-0403-8]
99. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis. Upper Saddle River: Prentice Hall; 2010.
100. Ziefle M, Röcker C. Acceptance of pervasive healthcare systems: A comparison of different implementation concepts. In: Proceedings of the 4th International Conference on Pervasive Computing Technologies for Healthcare. Acceptance of pervasive healthcare systems; 2010 Presented at: 4th International Conference on Pervasive Computing Technologies for Healthcare; 2010 March 22-25; Munich, Germany. [doi: 10.4108/icst.pervasivehealth2010.8915]
101. Sixsmith AJ. An evaluation of an intelligent home monitoring system. J Telemed Telecare 2000;6(2):63-72. [Medline: 10824373]
102. Townsend D, Knoefel F, Goubran R. Privacy versus autonomy: a tradeoff model for smart home monitoring technologies. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:4749-4752. [doi: 10.1109/IEMBS.2011.6091176] [Medline: 22255399]
103. Rowan J, Mynatt E. Digital Family Portrait Field Trial: Support for Aging in Place. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: ACM; 2005 Presented at: SIGCHI Conference on Human Factors in Computing Systems (CHI '05); 2005 April; Portland, Oregon, USA p. 521-530. [doi: 10.1145/1054972.1055044]
104. Courtney K, Demiris G, Rantz M, Skubic M. Needing smart home technologies: the perspectives of older adults in continuing care retirement communities. Journal of innovation in health informatics 2008 Nov 01;16(3):195-201. [doi: 10.14236/jhi.v16i3.694]
105. Lorenzen-Huber L, Boutain M, Camp LJ, Shankar K, Connelly KH. Privacy, Technology, and Aging: A Proposed Framework. Ageing International 2010 Dec 2;36(2):232-252. [doi: 10.1007/s12126-010-9083-y]

106. Luijkx K, Peek S, Wouters E. "Grandma, You Should Do It--It's Cool" Older Adults and the Role of Family Members in Their Acceptance of Technology. Int J Environ Res Public Health 2015 Dec 05;12(12):15470-15485 [FREE Full text] [doi: 10.3390/ijerph121214999] [Medline: 26690188]
107. Bandura A. Self-efficacy: the exercise of control. New York: W. H. Freeman; 1997.
108. Bandura A. Health promotion from the perspective of social cognitive theory. Psychology & Health 1998 Jul;13(4):623-649. [doi: 10.1080/08870449808407422]
109. Vorrink SNW, Antonietti AMGEF, Kort HSM, Troosters T, Zanen P, Lammers JJ. Technology use by older adults in the Netherlands and its associations with demographics and health outcomes. Assistive Technology 2016 Aug 22;29(4):188-196. [doi: 10.1080/10400435.2016.1219885]
110. Statistics Netherlands. CBS StatLine - Internet; toegang, gebruik en faciliteiten Internet. 2019. URL: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83429NED/table?dl=43EC [accessed 2020-07-17]
111. van Giesen RI, Fischer AR, van Dijk H, van Trijp HC. Affect and Cognition in Attitude Formation toward Familiar and Unfamiliar Attitude Objects. PLoS One 2015 Oct 30;10(10):e0141790 [FREE Full text] [doi: 10.1371/journal.pone.0141790] [Medline: 26517876]
112. Schwarz N. Attitude construction: evaluation in context. Soc Cogn 2007 Oct;25(5):638-656. [doi: 10.1521/soco.2007.25.5.638]
113. Fazio RH. Attitudes as object-evaluation associations of varying strength. Soc Cogn 2007 Oct 1;25(5):603-637 [FREE Full text] [doi: 10.1521/soco.2007.25.5.603] [Medline: 19424447]
114. Demiris G, Oliver DP, Dickey G, Skubic M, Rantz M. Findings from a participatory evaluation of a smart home application for older adults. THC 2008 May 01;16(2):111-118. [doi: 10.3233/thc-2008-16205]
115. Bossen C, Christensen LR, Grönvall E, Vestergaard LS. CareCoor: augmenting the coordination of cooperative home care work. Int J Med Inform 2013 May;82(5):e189-e199. [doi: 10.1016/j.ijmedinf.2012.10.005] [Medline: 23127539]

Abbreviations

AAL: ambient assisted living
AVE: average variance extracted
CFI: comparative fit index
EU: European Union
FIML: full information maximum likelihood
HTMT: heterotrait-monotrait
ICT: information and communication technology
RMSEA: root mean square error of approximation
SEM: structural equation modeling
SMC: squared multiple correlation
SRMR: standardized root mean square residual
TAM: Technology Acceptance Model
TLI: Tucker Lewis index
TPB: theory of planned behavior
UTAUT: unified theory of acceptance and use of technology

Edited by G Eysenbach; submitted 20.07.20; peer-reviewed by S Bhattacharjya, L Harst; comments to author 12.08.20; revised version received 06.10.20; accepted 17.01.21; published 31.03.21.

Please cite as:
Jaschinski C, Ben Allouch S, Peters O, Cachucho R, van Dijk JAGM
Acceptance of Technologies for Aging in Place: A Conceptual Model
J Med Internet Res 2021;23(3):e22613
URL: https://www.jmir.org/2021/3/e22613
doi:10.2196/22613
PMID:33787505

Review

Voice-Based Conversational Agents for the Prevention and Management of Chronic and Mental Health Conditions: Systematic Literature Review

Caterina Bérubé[1], MSc; Theresa Schachner[1], PhD; Roman Keller[2], MSc; Elgar Fleisch[1,2,3], Prof Dr; Florian v Wangenheim[1,2], Prof Dr; Filipe Barata[1], PhD; Tobias Kowatsch[1,2,3,4], PhD

[1]Center for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[2]Future Health Technologies Programme, Campus for Research Excellence and Technological Enterprise (CREATE), Singapore-ETH Centre, Singapore, Singapore

[3]Center for Digital Health Interventions, Institute of Technology Management, University of St. Gallen, St. Gallen, Switzerland

[4]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

Corresponding Author:
Caterina Bérubé, MSc
Center for Digital Health Interventions
Department of Management, Technology, and Economics
ETH Zurich
WEV G 214
Weinbergstrasse 56/58
Zurich, 8092
Switzerland
Phone: 41 44 633 8419
Email: berubec@ethz.ch

Abstract

Background: Chronic and mental health conditions are increasingly prevalent worldwide. As devices in our everyday lives offer more and more voice-based self-service, voice-based conversational agents (VCAs) have the potential to support the prevention and management of these conditions in a scalable manner. However, evidence on VCAs dedicated to the prevention and management of chronic and mental health conditions is unclear.

Objective: This study provides a better understanding of the current methods used in the evaluation of health interventions for the prevention and management of chronic and mental health conditions delivered through VCAs.

Methods: We conducted a systematic literature review using PubMed MEDLINE, Embase, PsycINFO, Scopus, and Web of Science databases. We included primary research involving the prevention or management of chronic or mental health conditions through a VCA and reporting an empirical evaluation of the system either in terms of system accuracy, technology acceptance, or both. A total of 2 independent reviewers conducted the screening and data extraction, and agreement between them was measured using Cohen kappa. A narrative approach was used to synthesize the selected records.

Results: Of 7170 prescreened papers, 12 met the inclusion criteria. All studies were nonexperimental. The VCAs provided behavioral support (n=5), health monitoring services (n=3), or both (n=4). The interventions were delivered via smartphones (n=5), tablets (n=2), or smart speakers (n=3). In 2 cases, no device was specified. A total of 3 VCAs targeted cancer, whereas 2 VCAs targeted diabetes and heart failure. The other VCAs targeted hearing impairment, asthma, Parkinson disease, dementia, autism, intellectual disability, and depression. The majority of the studies (n=7) assessed technology acceptance, but only few studies (n=3) used validated instruments. Half of the studies (n=6) reported either performance measures on speech recognition or on the ability of VCAs to respond to health-related queries. Only a minority of the studies (n=2) reported behavioral measures or a measure of attitudes toward intervention-targeted health behavior. Moreover, only a minority of studies (n=4) reported controlling for participants' previous experience with technology. Finally, risk bias varied markedly.

Conclusions: The heterogeneity in the methods, the limited number of studies identified, and the high risk of bias show that research on VCAs for chronic and mental health conditions is still in its infancy. Although the results of system accuracy and technology acceptance are encouraging, there is still a need to establish more conclusive evidence on the efficacy of VCAs for

the prevention and management of chronic and mental health conditions, both in absolute terms and in comparison with standard health care.

(J Med Internet Res 2021;23(3):e25933) doi:10.2196/25933

KEYWORDS

voice; speech; delivery of health care; noncommunicable diseases; conversational agents; mobile phone; smart speaker; monitoring; support; chronic disease; mental health; systematic literature review

Introduction

Background

Chronic and mental health conditions are increasingly prevalent worldwide. According to the World Health Statistics of 2020, noncommunicable diseases (eg, cardiovascular diseases, cancer, chronic respiratory diseases, and diabetes) and suicide are still the predominant causes of death in 2016 [1,2]. Although the underlying causes of these conditions are complex, behavior remains an important factor in their prevention and management. As the health care system is currently unfit to sustain the prevention and management of chronic and mental health conditions while containing its costs, continuous and personalized smartphone-based interventions have been developed to provide scaled-up behavioral support [3-6]. On the same note, conversational agents have been proven a valuable tool to deliver digital health interventions [7-9]. In particular, voice-based conversational agents (VCAs) have been shown to provide high user satisfaction in delivering interventions to influence healthy lifestyles [6].

VCAs can recognize human speech and, in turn, respond with synthesized speech. The human input is converted into an intent, triggering a specific information retrieval or function. This modality of interaction allows for hands-free access to some basic functions, such as searching for information on the internet, managing calendars, playing media content, calling, texting, emails, controlling internet-of-things devices and telling jokes [10,11]. Just as text-based [12,13] and embodied [14] conversational agents, VCAs have the potential to form an *alliance* [15] or *rapport* [16] with the patient through conversation, which is beneficial to treatment outcomes [17-19]. Compared with text-based interactions, however, voice-based interactions have several advantages. First, voice-based interaction leverages the naturalness [20,21] and social presence [22,23] of human-to-human conversation. Second, it facilitates input for users with low literacy or with visual [24], intellectual [25], motor, linguistic, and cognitive disabilities [26] and can support more natural health routine tasks when in-person health care is not possible [19,27]. Third, it opens the door to voice or speech analysis, whereas features of the patient's utterances can be passively monitored to derive health states [28-31]. Given the lack of agreement on the terminology [6], we will refer to VCAs to indicate the broad technology of dialog apps interacting with humans through speech recognition and synthesis.

VCAs are currently available on 2.5 billion devices worldwide, with smartphones being the leading type of devices, followed by smart speakers and computers. They can be found even in wearable technology, cars, and appliances [32,33]. Moreover, numerous health-related apps of VCAs are available [34]. Thus, these systems are increasingly used in our daily lives and are able to assist in the health care domain. In particular, commercial VCAs such as Amazon Alexa and Google Assistant are increasingly adopted and used as a framework by start-ups and health care organizations to develop products [35-40]. Although there is still room for improvement [41-43], curiosity in using VCAs for health care is growing. VCAs are used to retrieve health-related information (eg, symptoms, medication, nutrition, and health care facilities) [32,44]. This interest is even stronger in low-income households (ie, income <US $50,000 per year). Furthermore, when considering the accessibility of the voice modality for users with low literacy, VCAs could facilitate health management in countries where the education index is still relatively low [45] and smartphones are increasingly penetrating daily life [46] (eg, Brazil, Indonesia, Kenya, Mexico, Philippines, or South Africa).

To the best of our knowledge, only one scoping review has focused on VCAs for health [6]. The authors included research promoting self-management skills and healthy lifestyle behaviors in general and found that, although showing the feasibility of VCAs for health, the evidence was mostly preliminary. However, the authors do not inspect the methodology of the research in enough detail to define the methodological aspects that future research could improve. Thus, our contribution lies in a systematic review of VCA apps dedicated to the prevention and management of chronic and mental health conditions, which aims to provide a broader overview of the current state of research. Thus, we include evidence from both journals and conference papers and provide an overview of aspects affecting technology adoption, that is, system and user performance, ease of use, and attitude toward the target health behavior [47]. Furthermore, we highlight methodological aspects such as variables of interest, instruments used, population tested (in comparison with the target population), and VCA design description.

Objectives

This study aims to provide a better understanding of the current research on conversational agents delivering health interventions through voice-based interaction and to provide an overview of the methods and evaluations performed. We focus on VCAs specifically dedicated to the prevention and management of chronic and mental health conditions. As we focus on methods and findings in the domain of VCAs, comparing voice modality with others (eg, text and visual) is beyond the scope of this systematic literature review. Therefore, in this study, we seek to answer the following 2 questions: (1) What is the current evidence in favor of VCAs for the prevention and management of chronic and mental health conditions? (2) What are the methods used to evaluate them?

Methods

Reporting Standards

This study is compliant with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [48] (an overview of the study protocol is given in Multimedia Appendix 1 [49-55]).

Search Strategy

We conducted a systematic search of the literature available in July 2020 using the electronic databases PubMed MEDLINE, Embase, PsycINFO, Scopus, and Web of Science. These databases were chosen as they cover relevant aspects in the fields of medicine, technology, and interdisciplinary research and have also been used in other systematic reviews covering similar topics [7,8].

Search terms included items describing the constructs *voice modality*, *conversational agent*, and *health* (an overview of the search strategy is given in Multimedia Appendix 2).

Selection Criteria

We included studies if they (1) were primary research studies involving the prevention, treatment, or management of health conditions related to chronic diseases or mental disorders in patients; (2) involved a conversational agent; (3) the agent used voice as the main interaction modality; and (4) the study included either an empirical evaluation of the system in terms of system accuracy (eg, speech recognition and quality of answers), in terms of technology acceptance (eg, user experience, usability, likability, and engagement), or both.

Papers were excluded if they (1) involved any form of animation or visual representation, for example, embodied agents, virtual humans, or robots; (2) involved any form of health care service via telephone (eg, interactive voice response); (3) focused on testing a machine learning algorithm; and (4) did not target a specific patient population and chronic [49] or mental [50] health conditions.

We also excluded non-English papers, workshop papers, literature reviews, posters, PowerPoint presentations, and papers presented at doctoral colloquia. In addition, we excluded papers of which the authors could not access the full text.

Selection Process

All references were downloaded and inserted into a Microsoft Excel spreadsheet, and duplicates were removed. A total of 2 independent investigators conducted the screening for inclusion and exclusion criteria in 3 phases: first, we assessed the titles of the records; then their abstracts; and, finally, the full-text papers. After each of these phases, we calculated Cohen kappa to measure the inter-rater agreement between the 2 investigators. The interpretation of the Cohen kappa coefficient was based on the categories developed by Douglas Altman: 0.00-0.20 (poor), 0.21-0.40 (fair), 0.41-0.60 (moderate), 0.61-0.80 (good), and 0.81-1.00 (very good) [56,57]. The 2 raters consulted a third investigator in case of disagreements.

Data Extraction

A total of 2 investigators extracted data from the eligible papers into a Microsoft Excel spreadsheet with 52 columns containing information on the following aspects: (1) general information about the included papers, (2) voice-based interaction, (3) conversational agents, (4) targeted health conditions, (5) participants, (6) design, (7) measures, (8) main findings, and (9) additional study information such as funding information or conflicts of interest (a complete overview of the study characteristics is given in Multimedia Appendix 3 [52]).

We chose a narrative synthesis of the results and discussed and resolved any inconsistencies in the individual data extractions with a third investigator.

Risk of Methodological Bias

The choice of an appropriate risk of bias assessment tool was arbitrary, given the prevalence of conference papers and a wide variety of research designs in the included studies. Nevertheless, we wanted to evaluate the selected research concerning the transparency of reporting and the quality of the evidence. After extensive team discussions, the investigators decided to follow the approach of Maher et al [58], who devised a risk of bias assessment tool based on the CONSORT (Consolidated Standards of Reporting Trials) checklist [51]. The tool comprises 25 items and assigns scores of 0 or 1 to each item, indicating if the respective study satisfactorily met the criteria. Higher total scores indicated a lower risk of methodological bias. As the CONSORT checklist was originally developed for controlled trials and no such trials were included in our set of studies, we decided to exclude and adapt certain items as they were considered out of scope for this type of study. We excluded 3.b (*Trial design*), 6.b (*Outcomes*), 7.b (*Sample size*), 12.b (*Statistical methods*), and 14.b (*Recruitment*). Finally, item 17.b (*Outcomes and estimation*) was excluded and 17.a was fragmented into 2 subcriteria (ie, *Provides the estimated effect size* and *Provides precision*). A total of 2 investigators independently conducted the risk of bias assessment, and the differences were resolved in a consensus agreement (details are provided in Multimedia Appendix 4 [51,58]).

Results

Selection and Inclusion of Studies

In total, we screened 7170 deduplicated citations from electronic databases (Figure 1). Of these, we excluded 6910 papers during title screening. We further excluded 140 papers in the abstract screening process, which left us with 120 papers for full-text screening. After assessing the full texts, we found that 108 were not qualified. Cohen kappa was good in titles and full-text screening (κ=0.71 and κ=0.58, respectively), whereas it was moderate in abstract screening (κ=0.46). We explain the latter with a tendency of rater 1 to be more conservative than rater 2, giving a hypothetical probability of chance agreement of 50%. However, after meticulous discussion, the 2 investigators found a balanced agreement (an overview of the reasons for exclusion and the number of excluded records and Cohen kappa are shown in Figure 1) and considered 12 papers as qualified for inclusion and analysis (Table 1).

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

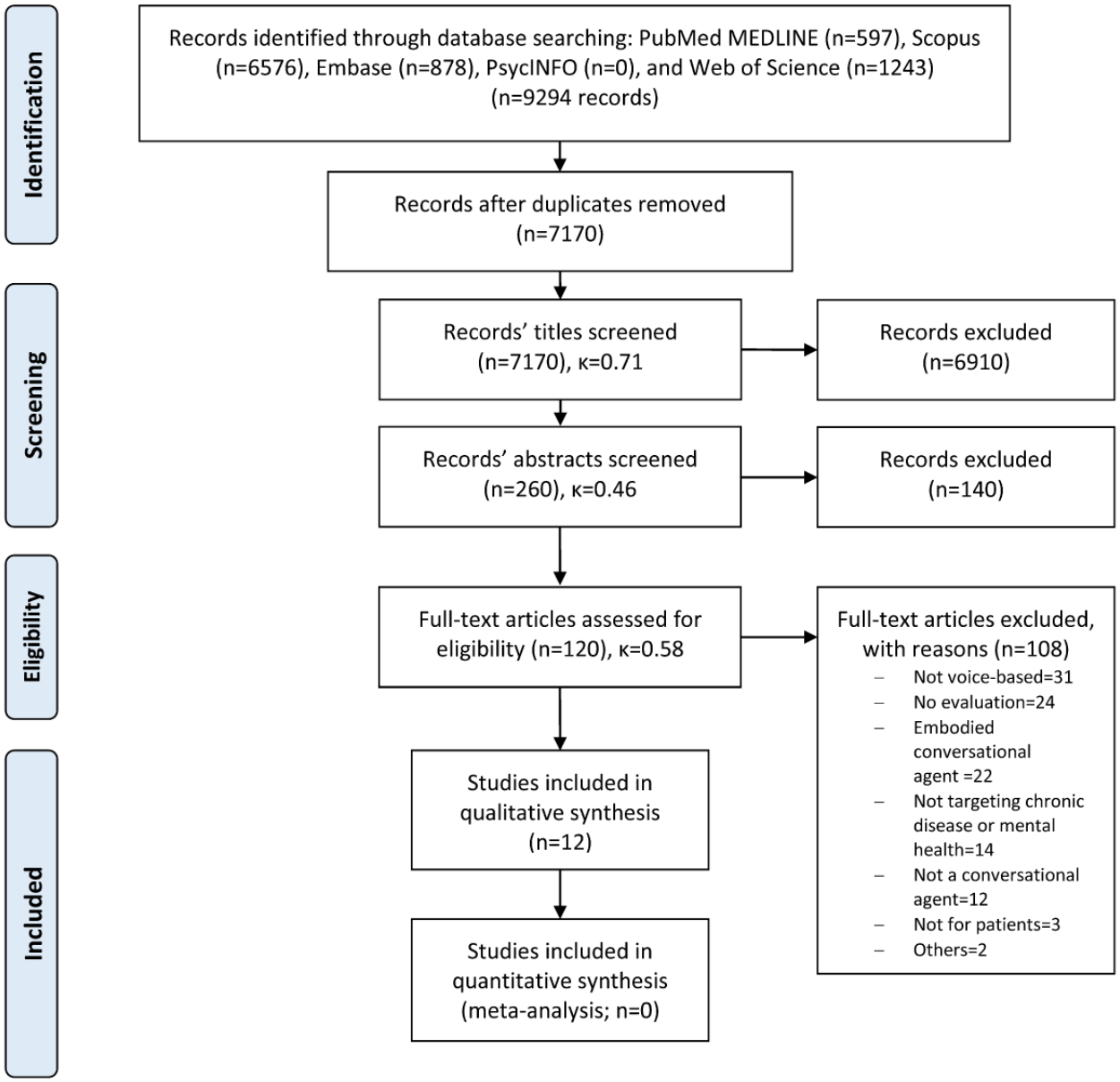


Table 1. Overview and characteristics of included records.

Reference, publication year	Study aim	Type of study participants	Addressed medical condition	Voice-enabled device type	Intervention category
Amith et al (2019) [59]	Development and acceptance evaluation	Healthy adults with at least one child under the age of 18 years (n=16)	Cancers associated with HPV[a]	Tablet	Support
Amith et al (2020) [60]	Development and acceptance evaluation	Healthy young adults aged between 18 and 26 years (n=24)	Cancers associated with HPV	Tablet	Support
Boyd and Wilson (2018) [61]	Criterion-based performance evaluation of commercial conversational agent	Authors as raters (n=2)	Cancers associated with smoking	Smartphone	Support
Cheng et al (2019) [62]	Development and acceptance evaluation	Older adults (n=10)	Diabetes (type 2)	Smart speaker	Monitoring and support
Galescu et al (2009) [63]	Development and performance evaluation	Chronic heart failure patients (n=14)	Heart failure	Not specified	Monitoring
Greuter and Balandin (2019) [64]	Development and performance evaluation	Adults with lifelong intellectual disability (n=9)	Intellectual disability	Smart speaker	Support
Ireland et al (2016) [65]	Development and acceptance evaluation	Adults recruited on campus (n=33)	Parkinson disease, dementia, and autism	Smartphone	Monitoring
Kadariya et al (2019) [66]	Development and acceptance evaluation	Clinicians and researchers (n=16)	Asthma	Smartphone	Monitoring and support
Lobo et al (2017) [67]	Development and acceptance evaluation	Healthy adults working regularly with senior patients (n=11)	Heart failure	Smartphone	Monitoring and support
Ooster et al (2019) [68]	Development and performance evaluation	Normal hearing (n=6)	Hearing impairment	Smart speaker	Monitoring
Rehman et al (2020) [69]	Development and performance and acceptance evaluation	Adults affiliated with the university (n=33)	Diabetes (type 1, type 2, gestational) and glaucoma	Smartphone	Monitoring and support
Reis et al (2018) [70]	Criterion-based performance evaluation of a commercial conversational agent	Not specified (n=Not specified)	Depression	Not specified	Support

[a]HPV: human papillomavirus.

Characteristics of the Included Studies

The publication years of the selected records ranged between 2009 and 2020, whereas the majority of the papers (n=5) were published in 2019. A total of 7 of the selected records were conference papers and 5 were journal papers.

The majority (n=10) of the selected papers developed and evaluated VCA [59,60,62-69], whereas 2 [61,70] aimed to report a criterion-based performance evaluation of existing commercial conversational agents (eg, Google Assistant and Apple Siri). Among the papers developing and evaluating a VCA, 6 [59,60,62,65-67] assessed the technology acceptance of the VCA, whereas 3 [63,64,68] assessed the system accuracy. Only one [69] assessed both performance and acceptance.

All studies (n=12) were nonexperimental [59-70], that is, they did not include any experimental manipulation. A total of 4 papers [61,66,68,70] did not explicitly specify the study design they used, whereas the other papers provided labels. One study stated conducting a *feasibility evaluation* [63], 1 a *focus group study* [65], 1 a *qualitative assessment of effectiveness and satisfaction* [62], and 1 a *case study* [69]. Furthermore, 1 conducted a *pilot study* [64], 2 declared deploying a *Wizard-of-Oz* (WOz) experiment [59,60], and 1 a *usability study* [67].

An overview of the included studies can be found in Table 1 (all details in Multimedia Appendix 3).

Main Findings

System Accuracy

Half (n=6) of the included studies [61,63,64,68-70] evaluated the accuracy of the system. In total, 4 of those studies [63,64,68,69] described precise speech recognition performance, whereas 3 [63,68,69] reported good or very good speech recognition performance, and 1 [64] study found mediocre recognition accuracy, with single-letter responses being slightly better recognized than word-based responses (details on speech recognition performance are given in Multimedia Appendix 5 [52]). A total of 2 studies [61,70] qualitatively assessed the accuracy of the VCAs. One study [61] observed that the standard Google Search performs better than a voice-activated internet search performed with Google Assistant and Apple Siri. Another study [70] reported on the accuracy of assisting with social

activities. They observed all commercial VCAs to perform well at basic greeting activities, Apple Siri and Amazon Alexa to perform the best in email management, and Apple Siri to perform the worst in supporting social games. Moreover, Google Assistant performed the best in social game activities but the worst in social media management.

Technology Acceptance

Of the 12 studies, 7 [59,60,62,65-67,69] reported technology acceptance findings, whereas the others (n=5) did not [61,63,64,68,70]. A total of 3 studies [60,66,67] reported technology acceptance through a System Usability Survey (SUS). One study [67] reported a relatively high usability score (SUS score of mean 88/100), whereas 1 study [60] described better usability of its VCA for human papillomavirus (HPV) in comparison with industry standards (ie, SUS score of mean 72/100). The latter also compared SUS scores between groups and found a higher score for participants who did not receive the HPV vaccine (mean 80/100), compared with those who did (mean 77/100) and the control group (mean 74/100). Note that the SDs of these results were not provided. In addition, the study found the score of Speech User Interface Service Quality to be medium (mean 4.29/7, SD 0.75). The third study [66] asked clinicians and researchers to evaluate the VCA with broader set of results. Clinicians and researchers rated the VCA with very good usability (ie, SUS score of mean 83.13/100 and 82.81/100, respectively) and very good naturalness (mean 8.25/10 and 8.63/10, respectively), information delivery (mean 8.56/10 and 8.44/10, respectively), interpretability (mean 8.25/10 and 8.69/10, respectively), and technology acceptance (mean 8.54/10 and 8.63/10, respectively). SDs of these results were not reported. A total of 2 studies [59,69] have reported different types of evaluations of technology acceptance. Thus, 1 study [59] reported good ease of use (mean 5.4/7, SD 1.59) and acceptable expected capabilities (mean 4.5/7, SD 1.46) but low efficiency (mean 3.3/7, SD 1.85) of its VCA, whereas the other [69] described a positive user experience of its VCA with all User Experience Questionnaire constructs. As the authors provided User Experience Questionnaire mean values per item we could only infer the mean values per construct manually. That is, Attractiveness mean score was 1.88/3; Perspicuity mean score was 1.93/3; Efficiency mean score was 1.88/3; Dependability mean score was 1.70/3; Stimulation mean score was 1.90/3; and Novelty mean score was 1.85/3. Note that the SDs of these results were not provided. Finally, 2 studies reported a qualitative evaluation of their VCA, one [62] stating theirs to be *more accepted than rejected* in terms of user satisfaction, without giving more details, and the other [65] mentioning a general *positive impression* but a slowness in the processing of their VCA.

Methodology of the Included Studies

We included all types of measures that were present in more than 1 study, that is, system accuracy measures, technology acceptance measures, behavioral measures, measures of attitude toward the target health behavior, and reported previous experience with technology.

The majority of the studies (n=10) did not report any behavioral measures [59-63,65-67,69,70], whereas 2 papers [64,68] did. One [68] described the frequency of verbal responses not relevant to the system (ie, nonmatrix-vocabulary words), whereas the other [64] provided engagement and user performance (ie, task completion, time to respond, points of difficulty, points of dropout, and quality of responses).

Half of the studies (n=6) did not report on any system measures [59,60,62,65-67], whereas the other half reported either speech recognition performance measures (n=4) [63,64,68,69] or criterion-based evaluation of the goodness of the VCA's response (n=2) [61,70]. In particular, 4 studies [63,64,68,69] measured speech recognition performance compared with human recognition. One of those [68] measured the accuracy of a diagnostic test score (ie, speech reception threshold) compared with the manually transcribed results. One study [64] measured the speech recognition percentage inferred from transcriptions of the interaction. One study [63] compared the VCA with nurse practitioners' interpretations of patients' responses. Finally, 1 [69] study gave more detailed results, reporting a confusion matrix; speech recognition accuracy, precision, sensitivity, specificity, and F-measure; and performance in task completion rate and prevention from security breaches.

Of the 12 studies, 7 [59,60,62,65-67,69] reported technology acceptance measures, whereas the remaining studies [61,63,64,68,70] did not. Although 2 studies [60,69] used validated questionnaires only and 2 [62,67] used adapted questionnaires only, 1 study used both validated and adapted questionnaires [66]. One study [59] used an adapted questionnaire and qualitative feedback as acceptance measures. One study [65] reported only qualitative feedback.

The majority of the included studies (n=10) did not provide measures of attitude toward the target health behavior [61-70]. The 2 remaining papers [59,60] provided validated questionnaires, and both focused on attitudes toward HPV vaccines. One study [59] used the Parent Attitudes about Childhood Vaccines, and 1 study [60] used the Carolina HPV Immunization Attitude and Belief Scale.

The majority of the included studies (n=8) also did not report controlling for participants' previous experience with technology [59-63,66,69,70]. Of the remaining 4 studies, 1 study [68] reported that all study participants had no experience with smart speakers; 1 [67] reported that all study participants were familiar with mobile health apps; and 1 [65] controlled for participants' smartphone ownership, use competence on Androids, iPhones, tablets, laptops, and desktop computers. Finally, 1 study [64] assessed the previous exposure of study participants to voice-based assistants but did not report on the results.

In general, risk bias varied markedly, from a minimum of 1 [70] to a maximum of 11.25 [60] and a median of 6.36 (more details are provided in Multimedia Appendix 4).

Health Characteristics

Of the included studies, cancer was the most common health condition targeted; 2 papers [59,60] addressed cancer associated with HPV, whereas 1 study [61] addressed cancer associated with smoking. The next most commonly addressed conditions were diabetes (n=2) [62,69] and heart failure (n=2) [63,67]. Other discussed conditions were hearing impairment [68],

asthma [66], and Parkinson disease [65]. A total of 3 papers addressed psychological conditions [64,65,70]. Specifically, they focused on dementia and autism [65], intellectual disability [64], and depression [70].

When inspecting the target population, we observed that 3 of the included studies [62,67,70] targeted older people, whereas 2 targeted either parents of adolescents [59] or pediatric patients [60]. The others targeted hearing-impaired individuals [68], smokers [61], patients with asthma [66], patients with glaucoma and diabetes [69], people with intellectual disability [64], and patients with chronic heart failure [63]. One study [65] did not specify a particular target population.

The actual study participants consisted of the following samples: healthy adults with at least one child under the age of 18 years (N=16) [59], healthy young adults aged between 18 and 26 years (N=24) [60], the authors themselves (N=2) [61], older adults (N=10) [62], patients with chronic heart failure (N=14) [63], adults with lifelong intellectual disability (N=9) [64], adults recruited on campus (N=33) [65], clinicians and researchers (N=16) [66], healthy adults working regularly with senior patients (N=11) [67], normal-hearing people (N=6) [68], and adults affiliated with a university (N=33) [69]. One study [70] did not specify the type or number of participants.

Characteristics of VCAs

A total of 8 studies [60,62,63,65-69] named their VCA, whereas 2 studies [59,64] did not specify any name (Multimedia Appendix 5). In total, 2 studies [61,70] did not provide a name because they evaluated existing commercially available VCA (ie, Amazon Alexa, Microsoft Cortana, Google Assistant, and Apple Siri).

The majority of the included studies (n=7) did not describe the user interface of their VCAs [60-62,64,68,70], whereas the remaining 5 papers did [59,65-67,69].

The underlying architecture of the investigated VCAs was described in 7 of the included studies [62,63,66-70], whereas 3 papers did not provide this information [61,64,65]. A total of 2 studies [59,60] could not provide any architectural information, given the nature of their study design (ie, WOz).

When considering the devices used to test the VCA, we found that smartphones were the most used (n=5) [61,65-67,69], followed by smart speakers (n=3) [62,64,68] and tablets (n=2) [59,60]. A total of 2 studies [63,70] did not specify which device they used for data collection.

The vast majority of the VCAs (n=10) were not commercially available [59,60,62-69] at the time of this systematic literature review. In particular, 1 study [65] reported the VCA to be available on Google Play store at the time of publication; however, the app could not be found by the authors of this literature review at the time of reporting (we controlled for geo-blocking by searching the app with an internet protocol address of the authors' country of affiliation [65]). Given that the other 2 studies tested consumer VCA, we classified these papers as testing commercially available VCAs [61,70].

Characteristics of Voice-Based Interventions

Interventions were categorized as either monitoring, support or both. Monitoring interventions refer to those focusing on health tracking (eg, symptoms and medication adherence), whereas support interventions include targeted or on-demand information or alerts. This categorization was based on the classification of digital health interventions by the World Health Organization [52]. A total of 5 VCAs [59-61,64,70] exclusively focused on support, and 3 studies [63,65,68] exclusively focused on monitoring. In total, 4 studies investigated a VCA providing both monitoring and support [62,66,67,69]. Monitoring activities were mainly implemented as active data capture and documentation (n=5) [62,63,66-69], whereas 1 study [66] also focused on self-monitoring of health or diagnostic data. One study [65] investigated self-monitoring of health or diagnostic data as the main monitoring activity.

Support services mainly consisted of delivering targeted health information based on health status (n=4) [59,60,64,67,69], whereas 1 study [67] also provided a lookup of health information. A total of 3 studies provided such a lookup of health information only [61,62,66], whereas 2 [62,66] also provided targeted alerts and reminders. Finally, 1 study delivered a support intervention in the form of task completion assistance [70] (more details on the interventions are given in Multimedia Appendix 3).

Discussion

Principal Findings

The goal of this study is to summarize the available research on VCA for the prevention and management of chronic and mental health conditions and provide an overview of the methodology used. Our investigation included 12 papers reporting studies on the development and evaluation of a VCA in terms of system accuracy and technology acceptance. System accuracy refers to the ability of the VCA to interact with the participants, either in terms of speech recognition performance or in terms of the ability to respond adequately to user queries. Technology acceptance refers to all measures of the user's perception of the system (eg, user experience, ease of use, and efficiency of interaction).

Most of the studies reported either one or the other aspect, whereas only 1 study reported both aspects. In particular, speech recognition in VCA prototypes was mostly good or very good. The only relevant flaw revealed was a slowness in the VCA responses, reported in 2 of the selected studies [59,65]. Commercial VCAs, although not outperforming Google Search when the intervention involved lookup of health information, seem to have a specialization in supporting certain social activities (eg, Apple Siri and Amazon Alexa for social media and office-related activities and Google Assistant for social games). These results suggest that there is great potential for noncommercial VCAs, as they perform well in the domain for which they were built, whereas commercial VCAs are rather superficial in their health-related support. Moreover, despite the heterogeneity of technology acceptance measures, the results showed good to very good performance. This suggests that the reviewed VCAs could satisfy users' expectations when

supporting the prevention and management of chronic or mental conditions. The evidence remains, however, hard to be conclusive. In fact, the majority of the included studies were published relatively recently, around 2019, and were fairly distributed between journal and conference or congress papers. Moreover, all studies were nonexperimental, and there was a general heterogeneity in the evaluation methods, especially in the user perception of the technology (ie, user experience). In particular, only 3 [60,66,69] of the 7 studies that included a measure of technology acceptance through a questionnaire [59,60,62,65-67,69] used a validated questionnaire, whereas the others adapted them. There was also a general discrepancy between the target population and the actual sample recruited. In particular, although the VCAs studied were dedicated to the management or prevention of chronic and mental health conditions, the evaluation was mainly conducted with healthy or convenience samples. Finally, according to our risk of bias assessment, the evidence is generally reported with insufficient transparency, leaving room for doubt about the generalizability of results, both in terms of technical accuracy and technology acceptance.

Considering the aforementioned aspects and the limited number of studies identified, it seems that research on VCAs for chronic diseases and mental health conditions is still in its infancy. Nevertheless, the results of almost all studies reporting system accuracy and technology acceptance are encouraging, especially for the developed VCAs, which inspires further development of this technology for the prevention and management of chronic and mental health conditions.

Related Work

To the best of our knowledge, this is the only systematic literature review addressing VCAs specifically dedicated to the prevention and management of chronic diseases and mental illnesses. Only 1 scoping review appraised existing evidence on voice assistants for health and focused on interventions of healthy lifestyle behaviors in general [6]. The authors highlight the importance of preventing and managing chronic diseases; however, although they report the preliminary state of evidence, they do not stress, for instance, specific methodological aspects that future research should focus on, to provide more conclusive evidence (eg, test on the actual target population). Moreover, the authors did not provide a measure of the preliminary state of evidence. However, it is important to inspect what aspects of the studies are most at risk of bias, to allow for a clearer interpretation of the results. Our review aims to highlight these aspects to provide meaningful evidence, not only for the scientific community in the field of disease prevention but also for this broad study population. We aimed to identify as precisely as possible the methodological gaps, to provide a solid base upon which future research can be crafted upon. For this reason, we first provide an overview of the instruments used and the variables of interest, distinguishing between behavioral and system and technology acceptance measures (compared with the sole outcome categorization), providing a more fine-grained overview of the methods used. Second, we provide a stronger argument in favor of the potential bias present in the research and, thus, the difficulty in interpreting the existing evidence, with a critical appraisal of the methodology, through a risk bias assessment. Moreover, the authors [6] included studies investigating the technology acceptance but excluded studies providing evidence on the technical performance of VCAs. However, this aspect has an important influence on the technology acceptance [71]. Thus, our review highlights the current state of research not only on the user's perception (ie, technology acceptance) but also on the device's ability to interact with the user (ie, technical performance). These aspects allowed us to provide a fair profile of the studies and to draw stronger conclusions on the methodology used to study a group of VCAs promoting the prevention and management of chronic diseases and mental illnesses.

Our findings are coherent with the review by Sezgin et al [6] in a series of aspects. First, we also show that research on VCAs is still emerging, with studies including small samples and focusing on the feasibility of dedicating VCA for a specific health domain. Second, we also find a heterogeneous set of target populations and target health domains. However, our findings are in contrast with those of Sezgin et al [6] in the following aspects. First, we report studies mainly focusing on developing and evaluating the system in terms of system accuracy or technology acceptance; Sezgin et al [6] also described efficacy tests but did not report on system accuracy. Third, the papers included in this study presented only VCA apps, whereas Sezgin et al [6] also included automated interventions via telephone. Finally, despite the preliminary character of the research, we include a risk bias assessment to formalize the importance of rigorous future research on VCAs for health.

In general, as we tried to include results explaining the technology acceptance of VCAs as a digital health intervention for the prevention and management of chronic and mental health conditions, our findings are more appropriate when concluding the current evidence-based VCAs in this specific domain rather than in healthy lifestyle behaviors in general.

Limitations

There are several limitations to our study, which may limit the generalizability of our results. First, our search strategy focused on nonspecific constructs (eg, health), which may have led to the initial inclusion of a large number of unrelated literature, in addition to that concerning the main topic of this review (ie, VCAs for chronic diseases and mental health). Given the infancy of this field, however, we chose a more inclusive strategy to avoid missing relevant literature for the analysis. Second, our systematic literature review aimed to assess the current scientific evidence in favor of VCAs for chronic diseases and mental health, thus not encompassing the developments of this technology in the industry. However, we aimed to summarize the findings and current methodologies used in the research domain and provided an overview of the scientific evidence on this technology. Third, to evaluate a possible experimental bias of the studies, we followed the reporting guidelines suggested by the *Journal of Medical Internet Research* and chose the CONSORT-EHEALTH checklist. Risk bias varied significantly among the selected studies. This evaluation scheme may be regarded as unsuitable for evaluating the presented literature, as none of the papers reported an experimental trial. An

evaluation scheme capable of taking into account the pioneering character of the papers concerning the use of this technology for health-related apps could have enabled a more differentiated assessment.

Future Work

The wide adoption of voice assistants worldwide and the interest in using them for health care purposes [32] have generated great potential for the effective implementation of scalable digital health interventions. There is, however, a lack of a clear implementation framework for VCAs. For instance, text-based and embodied conversational agents can currently be implemented using existing frameworks dedicated to digital health interventions [72-75]; however, to the best of our knowledge, there is no such framework for VCAs. A platform for the development of VCAs dedicated to specific chronic or mental health conditions could encourage standardized implementation, which would be more comparable in their development and evaluation processes. Currently, it is possible to develop apps for consumer voice assistants (eg, *skills* for Amazon Alexa or *actions* for Google Assistant). However, these products may be of privacy [76] or safety concerns [77]. Therefore, the academic community should strive for the creation of such a platform to foster the development of VCA for health.

The identified research provides diverse and general evaluation measures around technology acceptance (or user experience in general) and no evaluation based on theoretical models of health behavior (eg, intention of use). Thus, although the developed VCA might have been well received by the studied population samples, there is a need for a more systematic and comparable evaluation of the evidence systems to understand which aspects of VCAs are best for user satisfaction. Future research should favor the use of multiple standardized questionnaires dedicated to voice user interfaces [78] to further explore the factors potentially influencing their effectiveness (eg, rapport [79] and intention of use [71]).

This study reported the current state of research in the specific domain of VCAs for the prevention and management of chronic and mental health conditions in terms of behavioral, technological accuracy, and technology acceptance measures. However, the question remains as to how voice modality performs on these variables in comparison with other modalities, such as text-based conversational agents. Text-based conversational agents have been extensively studied in the domain of digital health interventions [80-83] and can be considered as a precursor to VCAs [9]. Moreover, voice modality may differ in their appropriateness of app, compared with text modality, depending on the health-related context (eg, public spaces [84,85] and type of user [24-26,86,87]). Thus, future research should not only standardize the research in terms of implementation and evaluation measures but also consistently evaluate this technology against what we could consider the gold standard of conversational agents.

Moreover, only 4 papers [63,64,68,69] compared the accuracy of the VCA's interpretation of participants' responses with humans' interpretation of participants' responses. Although it was limited to speech recognition, they were the only cases of human-machine comparison. To verify the suitability of VCAs as an effective and scalable complementary alternative to health care practitioners, more research should compare not only the system accuracy but also the general performance of this type of digital health intervention in comparison with standard in-person health care.

Finally, all papers conducted laboratory experiments and focused on short-term performance and technology acceptance. Even if this evidence shows the feasibility of VCAs for health care, it does not provide evidence on the actual effectiveness of VCAs in assisting patients in managing their chronic and mental health conditions compared with standard practices. Future research should provide evidence on complementary short-term and long-term measurements of technology acceptance and behavioral and health outcomes associated with the use of VCAs.

Conclusions

This study provides a systematic review of VCAs for the prevention and management of chronic and mental health conditions. Out of 7170 prescreened papers, we included and analyzed 12 papers reporting studies either on the development and evaluation of a VCA or on the criterion-based evaluation of commercial VCAs. We found that all studies were nonexperimental, and there was general heterogeneity in the evaluation methods. Considering the recent publication date of the included papers, we conclude that this field is still in its infancy. However, the results of almost all studies on the performance of the system and the experiences of users are encouraging. Even if the evidence provided in this study shows the feasibility of VCAs for health care, this research does not provide any insight into the actual effectiveness of VCAs in assisting patients in managing their chronic and mental health conditions. Future research should, therefore, especially focus on the investigation of health and behavioral outcomes, together with relevant technology acceptance outcomes associated with the use of VCAs. We hope to stimulate further research in this domain and to encourage the use of more standardized scientific methods to establish the appropriateness of VCAs in the prevention and management of chronic and mental health conditions.

Acknowledgments

This work was supported by the National Research Foundation, Prime Minister's Office, Singapore, under its *Campus for Research Excellence and Technological Enterprise Programme* and by the *CSS Insurance* (Switzerland).

Authors' Contributions

CB, EF, and TK were responsible for the study design and search strategy. CB and RK were responsible for the screening and data extraction. CB, RK, and TS were responsible for the data analysis. CB, RK, TS, and FB were responsible for the first draft. All authors were responsible for critical feedback and final revisions of the manuscript. TS and RK share second authorship. FB and TK share last authorship.

Conflicts of Interest

All authors are affiliated with the Center for Digital Health Interventions [88], a joint initiative of the Department of Management, Technology, and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurer CSS. EF and TK are also the cofounders of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, Pathmate Technologies was not involved in any way in the design, interpretation, analysis, or writing of the study.

Multimedia Appendix 1
Study protocol.
[PDF File (Adobe PDF File), 196 KB - jmir_v23i3e25933_app1.pdf]

Multimedia Appendix 2
Search terms per construct (syntax used in PubMed Medline).
[PDF File (Adobe PDF File), 98 KB - jmir_v23i3e25933_app2.pdf]

Multimedia Appendix 3
Complete list of characteristics of the included studies.
[PDF File (Adobe PDF File), 1588 KB - jmir_v23i3e25933_app3.pdf]

Multimedia Appendix 4
Risk-of-bias assessment.
[PDF File (Adobe PDF File), 946 KB - jmir_v23i3e25933_app4.pdf]

Multimedia Appendix 5
Main characteristics of the included studies.
[PDF File (Adobe PDF File), 115 KB - jmir_v23i3e25933_app5.pdf]

References

1. World Health Statistics 2020: monitoring health for the SDGs, sustainable development goals. Geneva: World Health Organization; 2020:1-77.
2. Suicide in the world: global health estimates. In: Document number: WHO/MSD/MER/19.3. Geneva: World Health Organization; 2019.
3. Kvedar JC, Fogel AL, Elenko E, Zohar D. Digital medicine's march on chronic disease. Nat Biotechnol 2016 Mar;34(3):239-246. [doi: 10.1038/nbt.3495] [Medline: 26963544]
4. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
5. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
6. Sezgin E, Militello L, Huang Y, Lin S. A scoping review of patient-facing, behavioral health interventions with voice assistant technology targeting self-management and healthy lifestyle behaviors. Transl Behav Med 2020 Aug 07;10(3):606-628. [doi: 10.1093/tbm/ibz141] [Medline: 32766865]
7. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
8. Schachner T, Keller R, Wangenheim VF. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. J Med Internet Res 2020 Sep 14;22(9) [FREE Full text] [doi: 10.2196/20701] [Medline: 32924957]
9. Car TL, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8) [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]

10. Ammari T, Kaye J, Tsai JY, Bentley F. Music, search, and IoT: how people (really) use voice assistants. ACM Trans Comput-Hum Interact 2019 Jun 06;26(3):1-28 [FREE Full text] [doi: 10.1145/3311956]
11. Hoy MB. Alexa, Siri, Cortana, and more: an introduction to voice assistants. Med Ref Serv Q 2018;37(1):81-88. [doi: 10.1080/02763869.2018.1404391] [Medline: 29327988]
12. Liu B, Sundar SS. Should machines express sympathy and empathy? Experiments with a health advice chatbot. Cyberpsychol Behav Soc Netw 2018 Oct;21(10):625-636. [doi: 10.1089/cyber.2018.0110] [Medline: 30334655]
13. Schwartzman CM, Boswell JF. A narrative review of alliance formation and outcome in text-based telepsychotherapy. Practice Innovations 2020 Jun;5(2):128-142. [doi: 10.1037/pri0000120]
14. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
15. Horvath AO, Luborsky L. The role of the therapeutic alliance in psychotherapy. J Consult Clin Psychol 1993;61(4):561-573. [doi: 10.1037/0022-006X.61.4.561]
16. Leach MJ. Rapport: a key to treatment success. Complement Ther Clin Pract 2005 Nov;11(4):262-265. [doi: 10.1016/j.ctcp.2005.05.005] [Medline: 16290897]
17. Mead N, Bower P. Patient-centred consultations and outcomes in primary care: a review of the literature. Patient Educ Couns 2002 Sep;48(1):51-61. [doi: 10.1016/s0738-3991(02)00099-x]
18. Martin DJ, Garske JP, Davis MK. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. J Consult Clin Psychol 2000;68(3):438-450. [doi: 10.1037/0022-006X.68.3.438]
19. Miner AS, Shah N, Bullock KD, Arnow BA, Bailenson J, Hancock J. Key considerations for incorporating conversational AI in psychotherapy. Front Psychiatry 2019;10:746 [FREE Full text] [doi: 10.3389/fpsyt.2019.00746] [Medline: 31681047]
20. Nass C, Steuer J, Tauber E. Computers are social actors. In: Proceedings of the Conference Companion on Human Factors in Computing Systems. 1994 Presented at: CHI '94: Conference Companion on Human Factors in Computing Systems; April, 1994; Boston Massachusetts USA. [doi: 10.1145/259963.260288]
21. Fogg B. Persuasive computers: perspectives and research directions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1998 Presented at: CHI98: ACM Conference on Human Factors and Computing Systems; April, 1998; Los Angeles California USA. [doi: 10.1145/274644.274677]
22. Cho E. Hey Google, can I ask you something in private? In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May, 2019; Glasgow Scotland UK. [doi: 10.1145/3290605.3300488]
23. Kim K, Norouzi N, Losekamp T, Bruder G, Anderson M, Welch G. Effects of patient care assistant embodiment and computer mediation on user experience. In: Proceedings of the IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). 2019 Presented at: 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR); Dec 9-11, 2019; San Diego, CA, USA. [doi: 10.1109/aivr46125.2019.00013]
24. Barata M, Salman GA, Faahakhododo I, Kanigoro B. Android based voice assistant for blind people. Library Hi Tech News 2018 Aug 06;35(6):9-11. [doi: 10.1108/lhtn-11-2017-0083]
25. Balasuriya SS, Sitbon L, Bayo AA, Hoogstrate M, Brereton M. Use of voice activated interfaces by people with intellectual disability. In: Proceedings of the 30th Australian Conference on Computer-Human Interaction. 2018 Presented at: OzCHI '18: 30th Australian Computer-Human Interaction Conference; December, 2018; Melbourne Australia.
26. Masina F, Orso V, Pluchino P, Dainese G, Volpato S, Nelini C, et al. Investigating the accessibility of voice assistants with impaired users: mixed methods study. J Med Internet Res 2020 Sep 25;22(9) [FREE Full text] [doi: 10.2196/18431] [Medline: 32975525]
27. Sezgin E, Huang Y, Ramtekkar U, Lin S. Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. NPJ Digit Med 2020;3:122 [FREE Full text] [doi: 10.1038/s41746-020-00332-0] [Medline: 33015374]
28. Pulido MLB, Hernández JBA, Ballester M, González CMT, Mekyska J, Smékal Z. Alzheimer's disease and automatic speech analysis: a review. Expert Syst Appl 2020 Jul;150. [doi: 10.1016/j.eswa.2020.113213]
29. Wang J, Zhang L, Liu T, Pan W, Hu B, Zhu T. Acoustic differences between healthy and depressed people: a cross-situation study. BMC Psychiatry 2019 Oct 15;19(1):300 [FREE Full text] [doi: 10.1186/s12888-019-2300-7] [Medline: 31615470]
30. Mota NB, Copelli M, Ribeiro S. Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance. NPJ Schizophr 2017;3:18 [FREE Full text] [doi: 10.1038/s41537-017-0019-3] [Medline: 28560264]
31. Tanaka H, Adachi H, Ukita N, Ikeda M, Kazui H, Kudo T, et al. Detecting dementia through interactive computer avatars. IEEE J Transl Eng Health Med 2017;5 [FREE Full text] [doi: 10.1109/JTEHM.2017.2752152] [Medline: 29018636]
32. Kinsella B, Mutchler A. Voice assistant consumer adoptoin in healthcare. 2019. URL: https://voicebot.ai/voice-assistant-consumer-adoption-report-for-healthcare-2019/ [accessed 2021-03-10]
33. Nearly half of Americans use digital voice assistants, mostly on their smartphones. Pew Research Center. 2017. URL: https://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/ [accessed 2021-03-10]
34. Chung AE, Griffin AC, Selezneva D, Gotz D. Health and fitness apps for hands-free voice-activated assistants: content analysis. JMIR Mhealth Uhealth 2018 Sep 24;6(9):e174 [FREE Full text] [doi: 10.2196/mhealth.9705] [Medline: 30249581]

35. Aiva: Virtual Health Assistant. URL: https://www.aivahealth.com/ [accessed 2020-11-23]
36. Orbita AI : leader in conversational AI for healthcare. Orbita. URL: https://orbita.ai/ [accessed 2020-11-23]
37. OMRON Health skill for Amazon Alexa. Omron. URL: https://omronhealthcare.com/alexa/ [accessed 2020-11-23]
38. Sugarpod. Wellpepper. URL: http://sugarpod.io/ [accessed 2020-11-23]
39. MFMER. Skills from mayo clinic. Mayo Foundation for Medical Education and Research. URL: https://www.mayoclinic.org/voice/apps [accessed 2020-11-23]
40. Guide your patients to the right care. Infermedica. URL: https://infermedica.com/ [accessed 2021-01-28]
41. López G, Quesada L, Guerrero L. Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces. In: Advances in Human Factors and Systems Interaction. Switzerland: Springer; 2017:A-50.
42. Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. JAMA Intern Med 2016 May 01;176(5):619-625 [FREE Full text] [doi: 10.1001/jamainternmed.2016.0400] [Medline: 26974260]
43. Miner AS, Milstein A, Hancock JT. Talking to machines about personal mental health problems. J Am Med Assoc 2017 Oct 03;318(13):1217-1218. [doi: 10.1001/jama.2017.14151] [Medline: 28973225]
44. Pradhan A, Lazar A, Findlater L. Use of intelligent voice assistants by older adults with low technology use. ACM Trans Comput-Hum Interact 2020 Sep 25;27(4):1-27. [doi: 10.1145/3373759]
45. Human Development Data (1990-2018). Human Development Reports : UNDP. URL: http://hdr.undp.org/en/data [accessed 2020-11-23]
46. Rosenberg S. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. 2019. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2019-02-05]
47. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
48. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. Br Med J 2015 Jan 02;350:g7647 [FREE Full text] [doi: 10.1136/bmj.g7647] [Medline: 25555855]
49. Fact sheets: chronic diseases. World Health Organization. URL: https://www.who.int/topics/chronic_diseases/factsheets/en/ [accessed 2021-03-10]
50. Fact sheets: mental disorders. World Health Organization. 2019. URL: https://www.who.int/news-room/fact-sheets/detail/mental-disorders [accessed 2021-03-10]
51. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. PLoS Med 2010 Mar 24;7(3) [FREE Full text] [doi: 10.1371/journal.pmed.1000251] [Medline: 20352064]
52. World Health Organization. Classification of digital health interventions v1.0. Sexual and reproductive health. 2018. URL: https://www.who.int/reproductivehealth/publications/mhealth/classification-digital-health-interventions/en/ [accessed 2021-03-12]
53. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Syst Rev 2012 Feb 09;1:2 [FREE Full text] [doi: 10.1186/2046-4053-1-2] [Medline: 22587842]
54. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
55. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. Br Med J 2015 Jan 02;350:g7647. [doi: 10.1136/bmj.g7647] [Medline: 25555855]
56. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 2016 Jul 02;20(1):37-46. [doi: 10.1177/001316446002000104]
57. Altman DG. Practical statistics for medical research. London, England: Chapman and Hall; 1990:1-624.
58. Maher CA, Lewis LK, Ferrar K, Marshall S, Bourdeaudhuij ID, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
59. Amith M, Zhu A, Cunningham R, Lin R, Savas L, Shay L, et al. Early usability assessment of a conversational agent for HPV vaccination. Stud Health Technol Inform 2019;257:17-23 [FREE Full text] [Medline: 30741166]
60. Amith M, Lin R, Cunningham R, Wu QL, Savas LS, Gong Y, et al. Examining potential usability and health beliefs among young adults using a conversational agent for HPV vaccine counseling. AMIA Jt Summits Transl Sci Proc 2020;2020:43-52 [FREE Full text] [Medline: 32477622]
61. Boyd M, Wilson N. Just ask Siri? A pilot study comparing smartphone digital assistants and laptop Google searches for smoking cessation advice. PLoS One 2018;13(3) [FREE Full text] [doi: 10.1371/journal.pone.0194811] [Medline: 29590168]

62. Cheng A, Raghavaraju V, Kanugo J, Handrianto Y, Shang Y. Development and evaluation of a healthy coping voice interface application using the Google home for elderly patients with type 2 diabetes. In: Proceedings of the 15th IEEE Annual Consumer Communications & Networking Conference (CCNC). 2018 Presented at: 15th IEEE Annual Consumer Communications & Networking Conference (CCNC); Jan 12-15, 2018; Las Vegas, NV, USA. [doi: 10.1109/ccnc.2018.8319283]
63. Galescu L, Allen J, Ferguson G, Quinn J, Swift M. Speech recognition in a dialog system for patient health monitoring. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshop. 2009 Presented at: IEEE International Conference on Bioinformatics and Biomedicine Workshop; Nov 1-4, 2009; Washington, DC, USA. [doi: 10.1109/bibmw.2009.5332111]
64. Greuter S, Balandin S, Watson J. Social games are funxploring social interactions on smart speaker platforms for people with disabilities. In: Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion. 2019 Presented at: CHI PLAY '19: The Annual Symposium on Computer-Human Interaction in Play; October, 2019; Barcelona Spain. [doi: 10.1145/3341215.3356308]
65. Ireland D, Atay C, Liddle J, Bradford D, Lee H, Rushin O, et al. Hello harlie: enabling speech monitoring through chat-bot conversations. Stud Health Technol Inform 2016;227:55-60. [Medline: 27440289]
66. Kadariya D, Venkataramanan R, Yip H, Kalra M, Thirunarayanan K, Sheth A. kBot: knowledge-enabled personalized chatbot for asthma self-management. In: Proceedings of the IEEE International Conference on Smart Computing (SMARTCOMP). 2019 Presented at: IEEE International Conference on Smart Computing (SMARTCOMP); June 12-15, 2019; Washington, DC, USA p. 138-143. [doi: 10.1109/smartcomp.2019.00043]
67. Lobo J, Ferreira L, Ferreira A. CARMIE: a conversational medication assistant for heart failure. Int J e-Health Med Commun 2017;8(4):1-17. [doi: 10.4018/ijehmc.2017100102]
68. Ooster J, Moreta PNP, Bach JH, Holube I, Meyer BT. “Computer, Test My Hearing”: accurate speech audiometry with smart speakers. In: ISCA Archive Interspeech 2019. 2019 Sep Presented at: Interspeech 2019; September 15-19, 2019; Graz URL: https://www.isca-speech.org/archive/Interspeech_2019/abstracts/2118.html [doi: 10.21437/interspeech.2019-2118]
69. Rehman UU, Chang DJ, Jung Y, Akhtar U, Razzaq MA, Lee S. Medical instructed real-time assistant for patient with glaucoma and diabetic conditions. Applied Sciences 2020 Mar 25;10(7):2216. [doi: 10.3390/app10072216]
70. Reis A, Paulino D, Paredes H, Barroso I, Monteiro M, Rodrigues V. Using intelligent personal assistants to assist the elderlies an evaluation of Amazon Alexa, Google Assistant, Microsoft Cortana, and Apple Siri. In: Proceedings of the 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW). 2018 Presented at: 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW); June 20-22, 2018; Thessaloniki, Greece. [doi: 10.1109/tishw.2018.8559503]
71. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]
72. MobileCoach. Center for Digital Health Interventions. URL: https://www.mobile-coach.eu/ [accessed 2021-03-10]
73. Filler A, Kowatsch T, Haug S, Wahle F, Staake T, Fleisch E. MobileCoach: A novel open source platform for the design of evidence-based, scalable and low-cost behavioral health interventions: Overview and preliminary evaluation in the public health context. In: Proceedings of the Wireless Telecommunications Symposium (WTS). 2015 Presented at: Wireless Telecommunications Symposium (WTS); April 15-17, 2015; New York, NY, USA. [doi: 10.1109/wts.2015.7117255]
74. Designing Conversational Agents for Healthcare and Beyond. Relational Agents Group. URL: http://relationalagents.com [accessed 2020-11-23]
75. Bickmore TW, Schulman D, Sidner CL. A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. J Biomed Inform 2011 Apr;44(2):183-197 [FREE Full text] [doi: 10.1016/j.jbi.2010.12.006] [Medline: 21220044]
76. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and Google Assistant. J Med Internet Res 2018 Sep 04;20(9) [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]
77. Chung H, Iorga M, Voas J, Lee S. Alexa, can I trust you? Computer (Long Beach Calif) 2017 Sep;50(9):100-104 [FREE Full text] [doi: 10.1109/MC.2017.3571053] [Medline: 29213147]
78. Kocaballi AB, Laranjo L, Coiera E. Measuring user experience in conversational interfaces: a comparison of six questionnaires. 32nd International BCS Human Computer Interaction Conference (HCI) 2018. [doi: 10.14236/ewic/hci2018.21]
79. Falkenström F, Hatcher RL, Skjulsvik T, Larsson MH, Holmqvist R. Development and validation of a 6-item working alliance questionnaire for repeated administrations during psychotherapy. Psychol Assess 2015 Mar;27(1):169-183. [doi: 10.1037/pas0000038] [Medline: 25346997]
80. Kowatsch T, Nißen M, Shih CHI, Rüegger D, Volland D, Filler A, et al. Text-based healthcare chatbots supporting patient and health professional teams: preliminary results of a randomized controlled trial on childhood obesity. Research Platform Alexandria. 2017. URL: https://www.alexandria.unisg.ch/publications/252944 [accessed 2021-03-10]

81. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. J Med Internet Res 2017 Jul 21;19(8):e267 [FREE Full text] [doi: 10.2196/jmir.7023] [Medline: 28784594]
82. Abd-Alrazaq A, Safi Z, Alajlani M, Warren J, Househ M, Denecke K. Technical metrics used to evaluate health care chatbots: scoping review. J Med Internet Res 2020 Jun 05;22(6) [FREE Full text] [doi: 10.2196/18301] [Medline: 32442157]
83. Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health – a scoping review. Verhaltenstherapie 2019 Aug 20:1-13. [doi: 10.1159/000501812]
84. Moorthy AE, Vu KPL. Voice activated personal assistant: acceptability of use in the public space. In: Information and Knowledge in Applications and Services. Switzerland: Springer; 2014:324-334.
85. Moorthy AE, Vu KL. Privacy concerns for use of voice activated personal assistant in the public space. Int J Huma-Compu Intera 2014 Dec 15;31(4):307-335. [doi: 10.1080/10447318.2014.986642]
86. Pradhan A, Mehta K, Findlater L. "Accessibility Came by Accident": use of voice-controlled intelligent personal assistants by people with disabilities. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI'18: CHI Conference on Human Factors in Computing Systems; April, 2008; Montreal QC Canada. [doi: 10.1145/3173574.3174033]
87. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Expert Syst Appl 2019 Sep;129:56-67 [FREE Full text] [doi: 10.1016/j.eswa.2019.03.054]
88. Center for Digital Health Interventions. URL: https://www.c4dhi.org [accessed 2021-03-24]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
HPV: human papillomavirus
SUS: System Usability Survey
VCA: voice-based conversational agent
WOz: Wizard-of-Oz

Edited by G Eysenbach; submitted 30.11.20; peer-reviewed by K Roberts, M Sobolev; comments to author 17.12.20; revised version received 10.02.21; accepted 03.03.21; published 29.03.21.

Please cite as:
Bérubé C, Schachner T, Keller R, Fleisch E, v Wangenheim F, Barata F, Kowatsch T
Voice-Based Conversational Agents for the Prevention and Management of Chronic and Mental Health Conditions: Systematic Literature Review
J Med Internet Res 2021;23(3):e25933
URL: https://www.jmir.org/2021/3/e25933
doi:10.2196/25933
PMID:33658174

Original Paper

Benchmarking Triage Capability of Symptom Checkers Against That of Medical Laypersons: Survey Study

Malte L Schmieding[1,2], MD; Rudolf Mörgeli[1], MD; Maike A L Schmieding[3]; Markus A Feufel[4*], Dipl-Ing (FH), MSc, PhD; Felix Balzer[1,2*], MSc, PhD, MD

[1]Department of Anesthesiology and Operative Intensive Care, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[2]Institute of Medical Informatics, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[3]Department of Biology, Chemistry, and Pharmacy, Institute of Pharmacy, Freie Universität Berlin, Berlin, Germany

[4]Department of Psychology and Ergonomics (IPA), Division of Ergonomics, Technische Universität Berlin, Berlin, Germany

[*]these authors contributed equally

Corresponding Author:
Felix Balzer, MSc, PhD, MD
Institute of Medical Informatics
Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 30 450 570425
Email: felix.balzer@charite.de

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/5/e30215

Abstract

Background: Symptom checkers (SCs) are tools developed to provide clinical decision support to laypersons. Apart from suggesting probable diagnoses, they commonly advise when users should seek care (*triage advice*). SCs have become increasingly popular despite prior studies rating their performance as mediocre. To date, it is unclear whether SCs can triage better than those who might choose to use them.

Objective: This study aims to compare triage accuracy between SCs and their potential users (ie, laypersons).

Methods: On Amazon Mechanical Turk, we recruited 91 adults from the United States who had no professional medical background. In a web-based survey, the participants evaluated 45 fictitious clinical case vignettes. Data for 15 SCs that had processed the same vignettes were obtained from a previous study. As main outcome measures, we assessed the accuracy of the triage assessments made by participants and SCs for each of the three triage levels (ie, *emergency care*, *nonemergency care*, *self-care)* and overall, the proportion of participants outperforming each SC in terms of accuracy, and the risk aversion of participants and SCs by comparing the proportion of cases that were overtriaged.

Results: The mean overall triage accuracy was similar for participants (60.9%, SD 6.8%; 95% CI 59.5%-62.3%) and SCs (58%, SD 12.8%). Most participants outperformed all but 5 SCs. On average, SCs more reliably detected emergencies (80.6%, SD 17.9%) than laypersons did (67.5%, SD 16.4%; 95% CI 64.1%-70.8%). Although both SCs and participants struggled with cases requiring self-care (the least urgent triage category), SCs more often wrongly classified these cases as emergencies (43/174, 24.7%) compared with laypersons (56/1365, 4.10%).

Conclusions: Most SCs had no greater triage capability than an average layperson, although the triage accuracy of the five best SCs was superior to the accuracy of most participants. SCs might improve early detection of emergencies but might also needlessly increase resource utilization in health care. Laypersons sometimes require support in deciding when to rely on self-care but it is in that very situation where SCs perform the worst. Further research is needed to determine how to best combine the strengths of humans and SCs.

(J Med Internet Res 2021;23(3):e24475) doi:10.2196/24475

KEYWORDS

digital health; triage; symptom checker; patient-centered care; eHealth apps; mobile phone; decision support systems; clinical; consumer health information; health literacy

Introduction

Use of Symptom Checkers

Patients obtain health-related information from health care professionals, but more frequently, information for patients is provided in print; on the web; and, most recently, via smartphone apps. Patients not only use these resources to supplement information received from health care professionals but also as a decision-support tool to advise them on whether and where to seek adequate health care, especially as health care pathways grow more complex. Symptom checkers (SCs) are tools developed to provide support to laypersons. Users can enter their complaints and, with some SCs, demographic or health-related information (eg, age, sex, and past medical history) to obtain advice on the urgency of their complaints (*triage advice*) and the most likely diagnosis. The demand for this type of support is evident; in the United States, 1 in 3 people reported resorting to the internet for self-diagnosis [1], and a study from 2019 found that half of the patients involved in that study had investigated their symptoms with an online search engine before going to an emergency department [2].

Evidence on SCs

Despite their popularity, there is no established framework to evaluate the performance of SCs [3,4]. The use of case vignettes, based on real or fictitious patients, has been a common approach for rating SCs [5-9]. The 2 most recent non–industry-funded audit studies using this methodology rated SC triage capability as unreliable, with an average of only 49% and 58% of appraisals deemed correct [10,11]. In line with these findings, a 2020 literature review concluded that most investigated SCs offered limited benefits [12].

A study showing that laypersons are just as capable of predicting criminal recidivism as a complex commercial algorithm [13] inspired us to compare the triage capability of SCs with that of participants with little or no medical training: are SCs merely a more complicated means of pointing out what an untrained individual could just as easily deduce? Is there an advantage to consulting SCs instead of relying on one's own judgment?

In addition to advising the individual user, SCs are also said to have the potential to reduce the burden on health care services. Unfortunately, not only has this potential benefit not materialized yet [3] but also there is evidence of the opposite effect, as overly risk-averse SCs promote more visits to emergency care services [14]. To address this issue, we also analyzed whether SCs were more risk averse than our participants. Although SCs can also provide diagnostic suggestions, we considered triage advice to be more relevant for assessing the impact of SC on use of health care resources and patient safety.

The purpose of this study is to benchmark the triage capability of SCs against that of their potential users, that is, laypersons.

Methods

Ethics Approval and Consent to Participate

This study was approved by the Ethics Committee of the Department of Psychology and Ergonomics (Institut für Psychologie und Arbeitswissenschaft) at Technische Universität Berlin (tracking number: FEU_03_20180615). Participants volunteered to participate in the survey, and informed consent was required.

Data Collection

This investigation builds on a prior study by Semigran et al [11], who evaluated SC triage performance based on case vignettes. We used their results on the performance of SCs as well as their case vignettes. Data were collected to determine the triage ability of medical laypersons, which was then used as a benchmark for comparing laypersons' performance with that of SCs.

Participants

All participants were US residents, at least 18 years of age, and had no professional medical background. Our investigation was limited to US residents, as the triage level definitions and the gold standard solutions assigned to the case vignettes by Semigran et al [11] might only be applicable to the US health care environment and might not apply to other health care systems with different service provider options.

Survey

We created an online survey with UNIPARK (QuestBack GmbH) [15] containing questions on demographics (age, sex, US residency, and highest level of completed formal education), past online searching behavior for medical information, 45 randomly ordered clinical case vignettes, and 5 attention checks (see *Procedure* for further details). We used the 45 case vignettes compiled and adjusted by Semigran et al [11], which are between 1 and 3 sentences long and describe a patient's signs and symptoms and occasionally mention elements of the patient's past medical history.

Participants were asked to classify each vignette into 1 of 3 triage categories, as defined by Semigran et al [11]: *emergency care*, involving "the advice to call an ambulance, go to an emergency department, or see a general practitioner immediately"; *nonemergency care*, which encompasses "advice to call a general practitioner or primary care provider, see a general practitioner or primary care provider, go to an urgent care facility, go to a specialist, go to a retail clinic, or have an e-visit"; and *self-care*, which is "advice to stay at home or go to a pharmacy." The definition of each triage level was explained at the beginning of the survey. The understanding of these

definitions by participants was ascertained by 3 control questions given before the case vignettes were presented. The questionnaire was piloted with 12 participants and refined according to their feedback to ensure readability and understandability.

Preparing the Case Vignettes

The 45 standardized case vignettes included 15 cases for each triage level. The vignettes, as chosen by Semigran et al [11], included both common and uncommon conditions with a wide range of chief complaints. The vignettes stemmed from various clinical sources, including material used to educate health care professionals.

For the purpose of our study, the vignettes were adapted to increase the comprehensibility of lay individuals. First, we transformed the bullet points into complete sentences. Second, we paraphrased technical terms. For example, we replaced "rhinorrhea" with "runny nose" and "tender" with "painful to the touch." In very few cases, explanations required elaboration. Our overall aim was to provide participants with the same information used by Semigran et al [11] to assess SCs. We deemed 1 case vignette vague regarding a crucial piece of information and had to supplement it with a detail left out in the Semigran et al [11] version of the vignette (see Multimedia Appendix 1 [11] for details). We retained the classification of the 45 case vignettes into 3 triage levels.

Understandability and paraphrasing were cross-validated by two native English speakers: one was a medical professional (RM) and the other was without a professional medical background (MALS). The adapted vignettes are shown in Multimedia Appendix 1.

Procedure

We recruited the participants through Amazon Web Service *Amazon Mechanical Turk* (MTurk), as it provides an established means to recruit US-based participants for sociopsychological surveys and is easy to access for researchers working outside of the United States [16]. Each participant received US $4.00 for completing the survey and a US $3.00 bonus if their overall accuracy in assigning the correct triage level was greater than or equal to 58%. The bonus was intended to provide an incentive for participants to pay close attention to the case vignettes and to assess a case's urgency as accurately as possible. The chosen threshold of 58% corresponds to outperforming the SC average reported by Semigran et al [11].

Two methods were employed to ensure that the participants paid close attention to the survey questions. First, we added 5 attention checks to the set of 45 case vignettes. These attention checks were formatted similarly to the case vignettes but included prompts to choose specific answer options. Participants were excluded from the analysis if they answered any of the 5 attention checks incorrectly. Second, upon completion of the survey, participants were asked to affirm that they were attentive and honest to improve the reliability of our data, as suggested in a reliability analysis on MTurk data [17]. We assured participants that they would be compensated for completing the survey even if they stated that they had responded inattentively or dishonestly. We analyzed data only from participants who affirmed their honesty and attentiveness.

The survey on MTurk was published on 3 different days (March 21, 2020, at 2 PM Pacific Daylight Time [PDT]; March 22, 2020, at 1:45 PM PDT; and March 29, 2020, at 1 PM PDT). By selecting the weekend day and early afternoon PDTs, we attempted to reach an MTurk population as diverse as possible, following a 2017 study on the intertemporal variation of the MTurk population [18]. On each day, participants were recruited within a few hours of publishing the survey.

Due to limited funding, the sample size was ultimately determined by the availability of funds and the number of participants who performed well enough to earn a bonus.

Data Analysis

Data were cleaned and explored using *R* 4.0.0 [19] and *tidyverse* packages [20]. Inferential analysis was conducted using the packages *lme4* [21] and *infer* [22]. Figures were created using the package *ggplot2* [23]. The data set containing participants' triage assessments and their demographic variables was made publicly available [24].

Following Semigran et al [11], we refer to each instance of an SC or a participant assessing a case vignette as a "case evaluation." For example, 2 participants each assessing all 45 case vignettes yielded 90 case evaluations.

Participant Characteristics

To assess the effects of demographic variables (age, sex, and educational level), a logistic regression was performed with the correct triage of a case vignette as a dependent variable. We calculated 95% CIs for the marginal probabilities of the fixed effects using the Wald method to assess whether demographic variables had a significant effect on participants' accuracy. The α level was set at .05.

Comparing SCs and Participants

For the comparison of SCs and participants, we performed (1) a comparison between participants and all rated SCs aggregated and (2) between participants and individual SCs.

Aggregate Comparison of SCs and Participants

The performance of the SCs was obtained from the appendix of the audit study by Semigran et al [11]. Comparisons were made between SCs and participants in terms of (1) triage accuracy, (2) tendency to overtriage (*risk aversion*), and (3) how difficult each case vignette was for the respective group (SCs and participants). Of the 15 SCs, 4 (*iTriage*, *Isabel*, *Symcat*, and *Symptomate*) were designed to never suggest self-care, with 1 SC (*iTriage*) always advising users to seek emergency care. To ensure that our results were not skewed by these special SCs, we conducted the main aggregate analyses twice, including and excluding those 4 SCs, and reporting results for both.

Triage Accuracy

Following Semigran et al [11], we compared the performance of SCs and participants at an aggregate level and for each triage level separately and overall. This was performed by calculating the sample's mean accuracy for SCs and participants, with

accuracy defined as the proportion of vignettes solved correctly. For the participants, the standard error of the sampling mean with 95% CIs was estimated by bootstrapping the participant data with 15,000 replications. The limits of the CI were calculated using the quantile method (2.5th and 97.5th quantile of the bootstrap sample means). The CIs for the SC sample were not calculated, as Semigran et al [11] sampled the SCs purposefully, that is, they selected which SCs to evaluate with care and not randomly.

Risk Aversion

The risk aversion of the SCs and the participants was determined using the ratio of overtriaged vignettes to undertriaged vignettes. We deemed a ratio greater than 1:1, which is more case vignettes overtriaged than undertriaged, as risk averse. To determine what type of triage mistakes were most likely to occur, we calculated the proportion of triage recommendations given in each triage category by SCs and by participants (eg, the proportion of evaluations in which participants recommended emergency care when self-care was appropriate or the proportion of evaluations in which SCs recommended nonemergency care when emergency care would have been the correct solution) and compared these proportions using the Pearson χ^2 test.

Difficulty of Case Vignettes

To analyze whether SCs and participants were challenged by the same case vignettes, the degree of difficulty of a case was calculated using the proportion of SCs and participants correctly triaging it. For example, if a case vignette was solved correctly by every SC, the vignette's degree of difficulty for SCs was 100%. SCs that did not evaluate the respective case vignette for technical reasons were not included in the denominator. A linear correlation analysis was then conducted to determine the relationship between case difficulty for SCs and case difficulty for participants.

Comparing Individual SCs With Participants

As users are likely to use only one or very few SCs, there is no basis for recommendations about using or not using SCs on an aggregated analysis alone. Therefore, additional analyses compared the performance of the participant group with each SC. Considering that most SCs did not evaluate every case vignette (due to technical reasons, see the study by Semigran et al [11]), the triage accuracy of the participants was calculated using only the cases evaluated by a specific SC, enabling a direct comparison. The CIs for participants' mean accuracy were calculated as described above. We also determined the proportion of participants that managed to achieve higher accuracy across cases than the respective SC. Furthermore, risk aversion was also evaluated, given the specific set of case vignettes for any given SC by plotting the proportion of vignettes that were overtriaged against the proportion of those undertriaged for participants versus SC.

Results

Participant Characteristics

Our survey was accessed 142 times in 3 days during which it was available in total, 51 participants were excluded, either for failing attention checks (n=41) or for not fulfilling the eligibility criteria (n=10). All the remaining participants affirmed that they had paid close attention during the survey and answered honestly. This yielded a total of 91 participants, each having assessed all 45 case vignettes, which totaled 4095 case evaluations by participants, 1365 for each triage level (Table 1).

The median time for completion of the survey (excluding the time for obtaining informed consent) was 20 minutes and 12 seconds (1st quartile=15 minutes:43 seconds; 3rd quartile=27 minutes:23 seconds). There was no significant difference in the participants' mean accuracy between the 3 sampling days. We detected no statistically significant influence of demographic variables on participants' triage accuracy.

Table 1. Participant characteristics (N=91).

Characteristics	Values
Age (years), median (range)	37 (20-73)
Gender, n (%)	
Female	36 (40)
Male	55 (60)
Education, n (%)	
Non–high school graduate	0 (0)
High school graduate	18 (20)
Some college	33 (36)
Bachelor's degree	36 (40)
Graduate degree	4 (4)
Recent[a] triage experience, n (%)	
Recently consulted an SC	20 (22)
Recently faced triage decision	23 (25)
Neither faced triage decision nor consulted an SC recently	62 (69)
Medical training, n (%)	
No training	80 (88)
Basic first aid training	11 (12)

[a]Recent was defined as "in the last 6 months."

Comparing SCs' and Participants' Triage Performance

Participant Performance

Overall, the participants triaged 3 out of 5 case vignettes correctly (2462/4065, 60.57%), and most participants qualified for the bonus payment (56/91, 62%). Their mean accuracy varied with triage level, roughly balanced for emergency and nonemergency situations (67.5% and 68.4%, respectively) but dropped below 50% for self-care vignettes. Of the 39.43% (1603/4065) of incorrect assessments, the majority (956/4065, 23.52%) were *overtriaged*, that is, participants assigned a more urgent triage level than necessary. Only about every sixth case vignette was *undertriaged* (647/4065, 15.92%), that is, participants assigned a less urgent triage level than necessary.

Aggregated Comparison Analyses

As most SCs were unable to evaluate at least one of the case vignettes, the 15 SCs assessing the 45 case vignettes yielded only 532 case evaluations (see the study by Semigran et al [11] for details): 183 for emergency vignettes, 175 for nonemergency vignettes, and 174 for self-care vignettes.

Triage Accuracy

At the aggregate level, SCs (58.0%; SD 12.8%) and participants (60.9%; SD 6.8%) showed very similar mean accuracies (Table 2). This remains to be the case when excluding the 4 SCs that did not suggest self-care (adjusted mean for the 11 SCs; 61.6%; SD 11.0%). Table 2 shows that differences become apparent when evaluating the triage levels separately: for emergency case vignettes, SCs outperformed the participants, whereas the participants outperformed the average SC in the nonemergency and self-care cases. For the least urgent triage level, this difference decreases when excluding those SCs that never recommend self-care.

Table 2. Mean triage accuracy of symptom checkers and participants.

Triage level	Percent triage accuracy, mean (SD)			95% CI
	All 15 SCs[a]	Subset of 11 SCs[b]	Participants[c]	
Emergency cases	80.6 (17.9)	79.8 (17.2)	67.5 (16.4)	64.1-70.8
Nonemergency cases	58.5 (29.1)	61.6 (27.8)	68.4 (13.8)	65.6-71.2
Self-care cases	30.6 (25.7)	41.8 (20.3)	46.7 (15.9)	43.4-49.8
Overall	58.0 (12.8)	61.6 (11.0)	60.9 (6.8)	59.5-62.3

[a]SC: symptom checker.

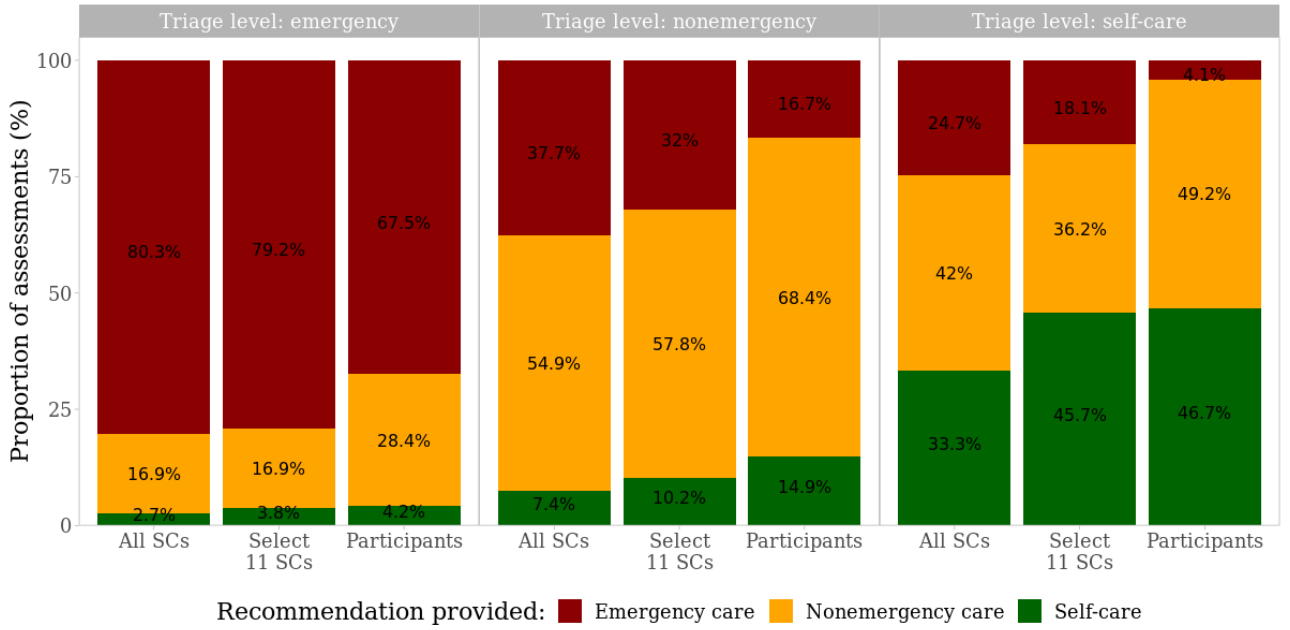
[b]For the subset of 11 SCs, SCs never recommending self-care or always recommending emergency care by design were excluded.

[c]For the participant sample, 95% CIs were calculated using bootstrapping.

Risk Aversion

The SCs were risk averse and overtriaged in more than a third of the evaluations (182/532, 34.2%), whereas undertriaging occurred in only 9.2% (49/532). Although participants also tended to be risk averse, this tendency was less pronounced (Figure 1). The ratio of overtriage to undertriage errors was 1.5:1 for participants whereas it was 3.5:1 for SCs. The SCs misclassified self-care cases as emergencies 6 times more often than participants did (43/174, 24.7% vs 56/1365, 4.10%) and 4.5 times more often (23/127, 18.1% vs 56/1365, 4.1%) when considering the subset of 11 SCs. The pair-wise differences in recommendations per triage level were statistically significant between participants and SCs (P=.002 for triage-level emergency [χ^2_2=12.5]; P<.001 for nonemergency [χ^2_2=46.3] and self-care [χ^2_2=109.6]). This holds true when comparing the participants' performance with the subset of 11 SCs (P=.02 for an emergency [χ^2_2=8.1] and P<.001 for a nonemergency [χ^2_2=19.0] and for self-care [χ^2_2=47.1]).

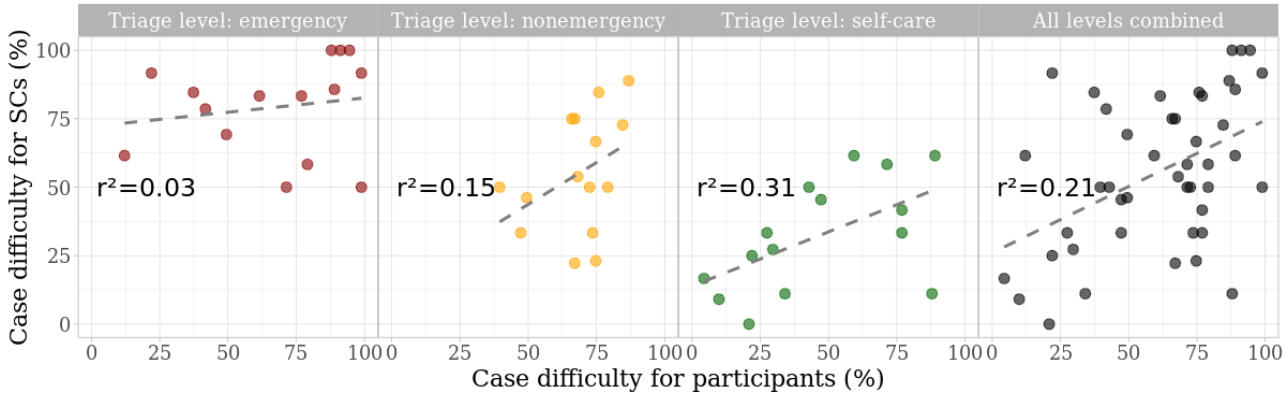
Figure 1. Triage evaluations by participants and SCs and triage level. "11 SCs" refers to the SC sample after exclusion of SCs that never recommend self-care (the least urgent triage level). SC: symptom checker.



Comparing Case Vignette Difficulty for SCs and for Participants

How challenging a case vignette was for SCs and participants varied widely: 3 vignettes were solved correctly by every SC and 1 vignette by none. Similarly, 4 vignettes were solved correctly by more than 90% of the participants and 2 by less than 10%. At every triage level, a broad variation in the degree of difficulty among case vignettes was observed. A very weak or no relationship could be detected for SCs and participants regarding case difficulty within each triage level (Figure 2).

Figure 2. Distribution of case difficulty for participants and SCs. Case difficulty is defined as the proportion of the group (SC or participants) evaluating the respective case correctly. The dashed line models a linear relationship. SC: symptom checker.



Comparing Individual SCs With Participants

As previously mentioned, an aggregated analysis of SCs is less meaningful than a direct comparison between the participant population and each SC, as users are likely to consult only one or very few SCs. The overall trend shows that the accuracy of both participants and SCs decreases for self-care vignettes (Figure 3).

A total of 5 SCs (*HMS* [Harvard Medical School] *Family Health Guide*, *Healthy Children*, *Steps2Care*, *Symptify*, and *Symptomate*) managed to outperform the participant sample, achieving an overall accuracy greater than the mean of the participants and its CI's upper limit (Table 3; see yellow dots in Figure 3). Five SCs had a triage capability lower than 80% (73/91) of the participants. This finding is partially explained by 3 of them apparently designed to never recommend self-care, hence failing in one-third of the cases owing to their design. One of these 3 SCs (*Isabel*) was outperformed only by a minority of participants (17/91, 18%), when self-care case vignettes were excluded from the analysis. The remaining 2 SCs (*Symcat* and *iTriage*) were still outperformed by most participants when self-care case vignettes were excluded. The participants' mean accuracy was stable at approximately 60%, independent of the slightly different samples of vignettes assessed by the SCs, with 2 exceptions: the participants were challenged by the sample of vignettes evaluated by *Healthy Children*, reaching a mean accuracy that was approximately 10% lower than in the other samples; conversely, the participants fared much better in assessing the vignette sample considered by *DoctorDiagnose*.

All but 2 SCs (*Family Doctor* and *Drugs.com*) were risk averse, making more overtriage errors than undertriage errors. Although the best 5 SCs were inclined toward overtriage, only one of them overtriaged more vignettes than the average participant (*Symptomate*; Figure 4).

Figure 3. Accuracy of SCs and participants by triage level (Em), nonemergency, and S-c. The accuracy of individual participants is indicated with blue dots. The aggregate accuracies of participants are shown as box plots. Em: emergency; SC: symptom checker; S-c: self-care.

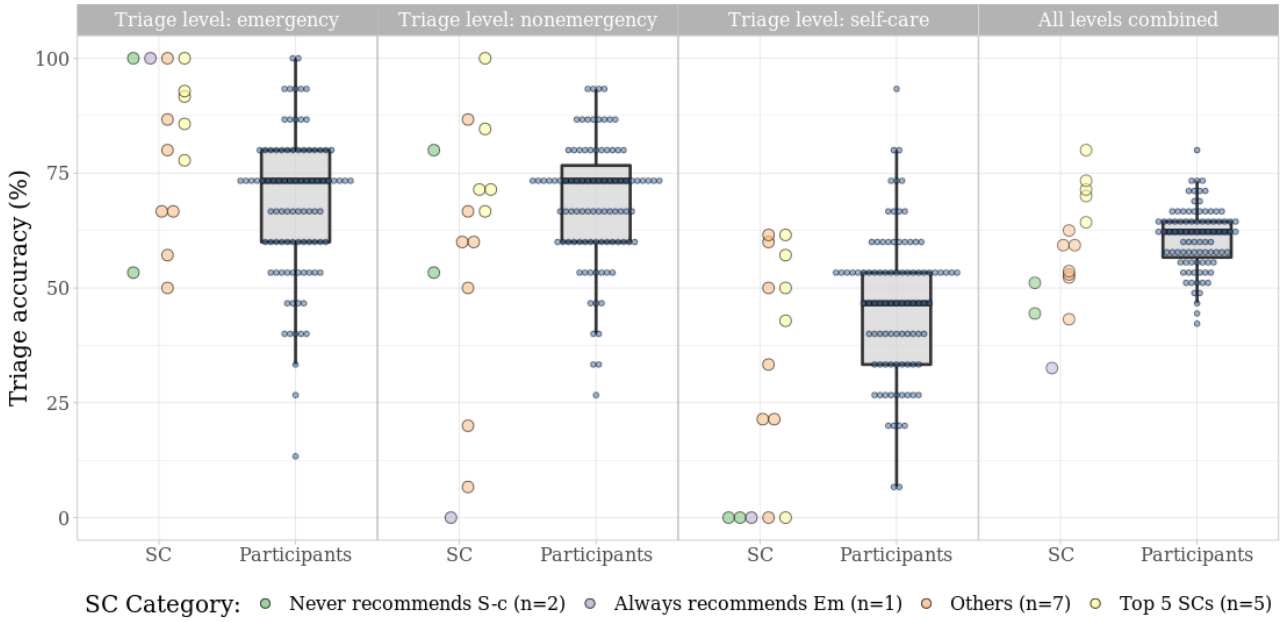


Table 3. Comparison of accuracy between symptom checkers and participants.

SC[a,b] name	Accuracy[c], n (%)	Participants		Comparison
		Percent accuracy[d,e], mean (SD)	95% CI	Percentage of participants outperforming the SC (95% CI)[d,e]
HMS[f] Family Health Guide, n=40	32 (80)	59.5 (7.1)	58.0-60.9	0 (0-0)
Healthy Children, n=15	11 (73)	49.9 (10.1)	47.7-52.1	1.1 (0-3.3)
Steps2Care, n=42	30 (71)	59.7 (7.2)	58.2-61.1	1.1 (0-3.3)
Symptify, n=40	28 (70)	60.2 (7.2)	58.2-61.7	5.5 (1.1-11.0)
Symptomate[g], n=14	9 (64)	60.9 (11.6)	58.6-63.2	26.4 (17.6-35.2)
Drugs.com, n=42	25 (59)	60.6 (6.5)	59.3-61.9	51.6 (41.8-61.5)
FreeMD, n=44	26 (59)	60.2 (6.7)	58.9-61.6	56.0 (45.1-65.9)
Doctor Diagnose, n=16	10 (62)	69.5 (10.9)	67.3-71.7	63.7 (53.8-73.6)
Family Doctor, n=41	22 (53)	58.1 (7.0)	56.7-59.6	68.1 (58.2-78.0)
Early Doc, n=17	9 (52)	63.4 (11.4)	61.1-65.7	76.9 (68.1-85.7)
Isabel[g], n=45	23 (51)	60.9 (6.8)	59.4-62.2	89 (82.4-94.5)
NHS[h], n=44	23 (52)	62.0 (6.9)	60.9-63.4	89 (82.4-94.5)
Symcat[g], n=45	20 (44)	60.9 (6.8)	59.5-62.2	97.8 (94.5-100)
Healthwise, n=44	19 (43)	61.2 (7)	59.7-62.6	98.9 (96.7-100)
iTriage[h,i], n=43	14 (32)	60.5 (6.9)	59.1-61.9	100 (100-100)

[a]SC: symptom checkers

[b]SCs are listed in order by the proportion of participants outperforming them.

[c]Most SCs did not evaluate every case vignette. Their accuracy is given as the proportion of correctly solved vignettes of the total vignettes that they evaluated.

[d]The participants' accuracy is based on their assessment of the same case vignettes assessed by the respective SC.

[e]For the participant sample, 95% CIs were calculated using bootstrapping.

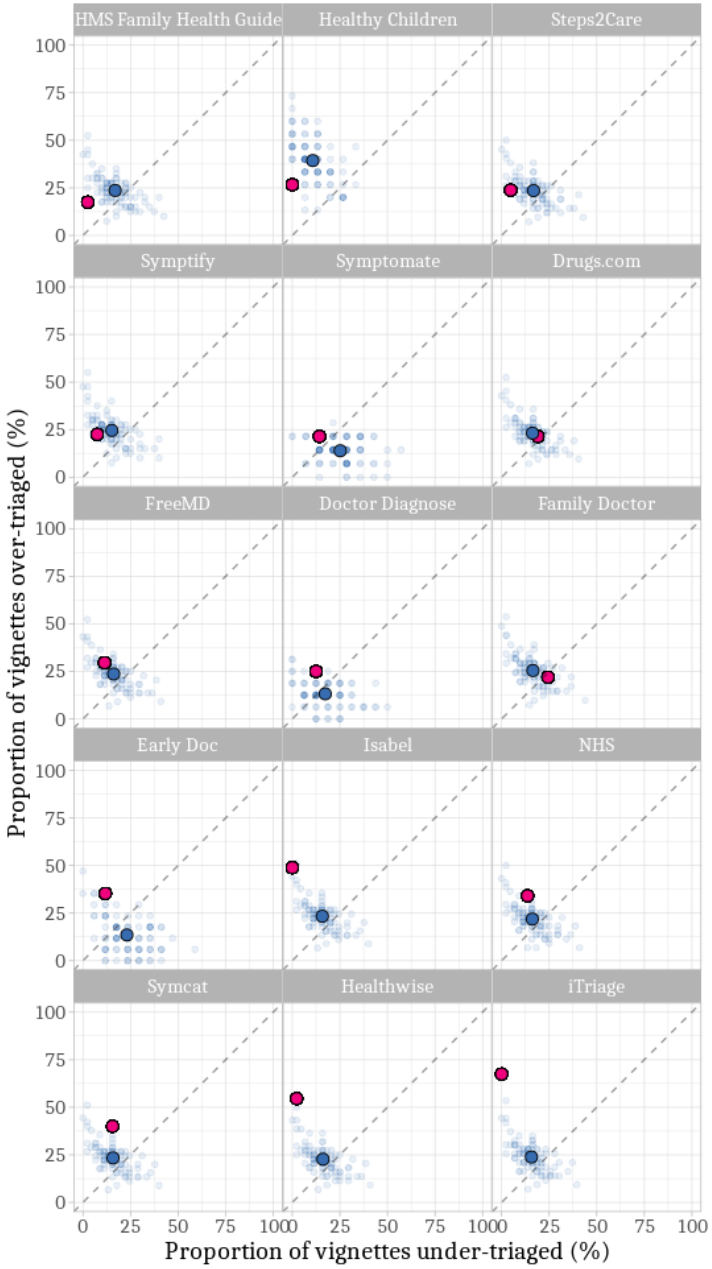
[f]HMS: Harvard Medical School.

[g]Four SCs were apparently designed never to recommend self-care.

[h]NHS: National Health Service.

[i]One SC advised seeking emergency care for all case vignettes.

Figure 4. Comparison of the overtriage inclination of symptom checkers (SCs) and participants. The dashed line shows where proportions of over and undertriaged errors are equal. Proximity to the left lower corner indicates a high triage accuracy. The red dot marks the respective symptom checker. The faded blue dots refer to the performance of individual participants. The larger blue dot marks their average performance. The SCs are ordered from left to right and top to bottom by the proportion of participants outperforming them, with the lowest proportional difference at the top left and the highest proportional difference on the bottom right.



Discussion

Principal Findings

Our study suggests that an average SC has no greater overall triage accuracy than an average user. However, this does not imply that SCs are not useful. Specifically, our data confirm a prior study showing that the lay population has difficulties reliably identifying medical emergencies [25]. On average, participants failed to identify every third emergency, and 12% (11/91) of our participants identified emergencies less reliably than the worst-performing SC.

Most SCs tended to overtriage. From a clinical and legal perspective, it can make sense to accept the resulting inflated cost of false alarms to avoid potentially missing an emergency (*defensive decision making*). In contrast, false alarms raised by SCs can functionally exacerbate overcrowding in health care services. In fact, the ability of some SCs to reliably detect emergencies can be partially attributed to their general tendency—by design—to recommend emergency care even for self-care cases (the least urgent triage level) where no medical care is warranted. This trade-off must be considered before recommending their use.

Studies on the effects of SC advice on users are scarce. Therefore, general recommendations on whether laypersons should use SCs cannot be formulated as yet. On the basis of a detailed analysis of the performance variation among SCs and human decision makers, we showed that the five best SCs that Semigran et al [11] included in their sample outperformed almost all our participants and thus could be seen as beneficial to users. In contrast, SCs mistake self-care cases for emergencies a substantial number of times. This hints at SCs being better suited to help users who are looking for an answer on where they should seek professional help (ie, by discriminating between emergency and nonemergency cases) rather than on whether they should seek medical care at all (ie, by discriminating between self-care and non–self-care cases).

Finally, SCs and participants struggled with different kinds of case vignettes, that is, SCs performed poorly in some clinical situations, whereas in others, their performance was superior to that of their users. For example, the 15 pediatric cases evaluated by the SC *Healthy Children* appear to have been more challenging for participants (mean accuracy of 49.9%) than the 30 nonpediatric cases (mean accuracy of 66.3%). To provide a more differentiated picture of SC triage performance, further analyses should also investigate performance differences with respect to different types of cases.

Limitations

Compared with the general population of the United States [26], our participants were better educated and included more men than women. The median and mean ages were similar to those of the general US population. One study suggests that the groups most likely to seek health information online are younger White females from high-income households, most with a bachelor's degree or higher [1]. Most participants in a survey among users of a specific SC (Isabel) were female and White but older than the average population [27]. Despite the fact that our sample's demographic distribution did not fully resemble the US population or, presumably, the population of SC users, we consider our findings to have at least some external validity for these populations, as demographic variables showed no significant influence on triage accuracy.

The data on SCs date back to a study published in 2015 [11], where the specific versions of the SCs assessed were not specified. Therefore, changes in performance due to possible upgrades were not considered. Such upgrades are likely, and new SCs have since entered the market. Other SCs included in the Semigran et al sample [11] are no longer available online, including the best-performing SC (*HMS Family Health Guide*). This speaks to the general problem that future research evaluating the performance of SCs will have to address the rapidly changing markets and technological developments.

As we built our study on the materials of the Semigran et al study [11], we also inherited their limitations: the chosen 45 case vignettes do not cover the entire spectrum of prehospital case presentations, especially with the omission of mental health–related scenarios. In addition, some case vignettes lacked a proper diagnosis and stated only the presenting complaints (eg, "Vomiting" for vignette 45, "Constipation" for vignette 40, "Back pain" for vignette 20). This prevented a plausibility check of the gold standard triage level that should be assigned to each vignette.

In general, assessing triage capability with case vignettes has limited validity. This limitation is arguably greater for human participants than for SCs. Although SCs assess a case with a set algorithm and are therefore dependent only on input, contextual (social, emotional, etc) factors play a greater role in human decision making. In a real-life setting, humans might also notice and process more or less information than presented in a case vignette. In addition, reading "back pain" in a dry case vignette is surely a different matter than experiencing it. Thus, our results might be more valid for situations where SC users utilize the tool to triage someone other than themselves. Research shows that this is common practice, as up to 50% of online health information seekers do so on behalf of someone else [1].

Conclusions

Prior publications have emphasized the need for a framework within which the safety and usefulness of SCs should be analyzed. Assessing the average performance of SCs, as has often been done, fosters few actionable recommendations. Given the high-performance variability among SCs, we consider benchmarking with case vignettes as a valuable first step in identifying the best SCs, which could then be tested extensively against relevant competitors.

Although comparing SCs' triage capability against that of health care professionals is certainly useful [28], this approach implicitly asks whether the former could replace the latter, rather than assessing whether and under which circumstances a user should rely on an SC or refrain from using it. Similar to the common practice of testing a new medicine against a placebo, we suggest that SCs should be benchmarked against a realistic alternative, for example, an SC user relying on his own appraisal (stand-alone triage capability).

Following this approach, our study suggests that the lay population would benefit from some SCs to some extent. Although SCs detect emergencies more reliably than the average user, they are more risk averse than the general population and recommend emergency care more often than is actually necessary. This is a cause for concern, as it might unnecessarily increase the burden on already overwhelmed health care services. Thus, advice on when not to seek emergency care would be the most useful feature of SCs, but it is precisely in that situation that they performed the worst. Further research should investigate which user groups benefit the most from using SCs and whether it is possible to identify the characteristics of scenarios where laypersons are superior to SCs in assessing triage levels. The detailed analyses presented in this paper provide a first step toward a framework for comparatively assessing the respective weaknesses and strengths of both SCs and human decision makers to be able to determine when humans should rely on SCs rather than on their gut feeling and vice versa.

Acknowledgments

The authors express their gratitude to the participants, to Felix Grün for his support in designing the questionnaire and for his valuable feedback, to Eike Richter for his advice on statistical methods, and to Frances Lorié for proofreading the manuscript. The project was funded by the home institutions of the previous authors (MF and FB). No external funding was required for this study. The authors acknowledge support from the German Research Foundation (DFG) and the Open Access Publication Fund of Charité—Universitätsmedizin Berlin.

Authors' Contributions

MS conceived the study, created the questionnaire, designed and conducted the analyses, and wrote the first draft of the paper. MALS assisted with case vignette adaptations. RM assisted with case vignette adaptations and manuscript development. FB and MF provided critical input and advised on the study and questionnaire design, analysis methods, and drafts of the paper. FB and MF contributed equally and share the last authorship. All authors accept full responsibility for the final version of the paper. The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Conflicts of Interest

All authors have completed the International Committee of Medical Journal Editors uniform disclosure form and declare no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

Multimedia Appendix 1
Adapted case vignettes and case difficulty level.
[DOCX File , 40 KB - jmir_v23i3e24475_app1.docx]

References

1. Fox S, Duggan M. Health online 2013. Pew Research Center. 2013. URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2021-02-18]
2. Martin SS, Quaye E, Schultz S, Fashanu OE, Wang J, Saheed MO, et al. A randomized controlled trial of online symptom searching to inform patient generated differential diagnoses. NPJ Digit Med 2019 Nov 11;2(1):110 [FREE Full text] [doi: 10.1038/s41746-019-0183-0] [Medline: 31728417]
3. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 01;9(8) [FREE Full text] [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
4. Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. The Lancet 2018 Nov;392(10161):2263-2264. [doi: 10.1016/s0140-6736(18)32819-8]
5. Bavdekar SB, Pawar M. Evaluation of an Internet-Delivered Pediatric Diagnosis Support System (ISABEL®) in a Tertiary Care Center in India. Indian Pediatrics. 2005. URL: http://www.indianpediatrics.net/nov2005/1086.pdf [accessed 2021-02-18]
6. Berner ES, Webster GD, Shugerman AA, Jackson JR, Algina J, Baker AL, et al. Performance of Four Computer-Based Diagnostic Systems. N Engl J Med 1994 Jun 23;330(25):1792-1796. [doi: 10.1056/nejm199406233302506]
7. Bond WF, Schwartz LM, Weaver KR, Levick D, Giuliano M, Graber ML. Differential diagnosis generators: an evaluation of currently available computer programs. J Gen Intern Med 2012 Feb 26;27(2):213-219 [FREE Full text] [doi: 10.1007/s11606-011-1804-8] [Medline: 21789717]
8. Farmer N. An update and further testing of a knowledge-based diagnostic clinical decision support system for musculoskeletal disorders of the shoulder for use in a primary care setting. J Eval Clin Pract 2014 Oct 15;20(5):589-595. [doi: 10.1111/jep.12153] [Medline: 24828447]
9. Farmer N, Schilstra MJ. A Knowledge-based Diagnostic Clinical Decision Support System for Musculoskeletal Disorders of the Shoulder for Use in a Primary Care Setting. Shoulder & Elbow 2017 Feb 06;4(2):141-151. [doi: 10.1111/j.1758-5740.2011.00165.x]
10. Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Med J Aust 2021 Feb 09;214(3):143. [doi: 10.5694/mja2.50923] [Medline: 33423279]
11. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. Br Med J 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
12. Gottliebsen K, Petersson G. Limited evidence of benefits of patient operated intelligent primary care triage tools: findings of a literature review. BMJ Health Care Inform 2020 May 07;27(1):e100114. [doi: 10.1136/bmjhci-2019-100114]

13. Dressel J, Farid H. The accuracy, fairness, and limits of predicting recidivism. Sci Adv 2018 Jan 17;4(1):eaao5580 [FREE Full text] [doi: 10.1126/sciadv.aao5580] [Medline: 29376122]
14. Anhang Price R, Fagbuyi D, Harris R, Hanfling D, Place F, Taylor TB, et al. Feasibility of web-based self-triage by parents of children with influenza-like illness: a cautionary tale. JAMA Pediatr 2013 Feb 01;167(2):112-118. [doi: 10.1001/jamapediatrics.2013.1573] [Medline: 23254373]
15. Unipark. QuestBack. URL: https://www.unipark.com/ [accessed 2021-02-18]
16. Mortensen K, Hughes TL. Comparing Amazon's Mechanical Turk Platform to Conventional Data Collection Methods in the Health and Medical Research Literature. J Gen Intern Med 2018 Apr 4;33(4):533-538 [FREE Full text] [doi: 10.1007/s11606-017-4246-0] [Medline: 29302882]
17. Rouse SV. A reliability analysis of Mechanical Turk data. Computers in Human Behavior 2015 Feb;43:304-307. [doi: 10.1016/j.chb.2014.11.004]
18. Casey LS, Chandler J, Levine AS, Proctor A, Strolovitch DZ. Intertemporal differences among MTurk workers: time-based sample variations and implications for online data collection. SAGE Open 2017 Jun 14;7(2). [doi: 10.1177/2158244017712774]
19. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing URL: https://www.R-project.org/ [accessed 2021-02-18]
20. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw 2019 Nov;4(43):1686. [doi: 10.21105/joss.01686]
21. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using. J Stat Soft 2015;67(1). [doi: 10.18637/jss.v067.i01]
22. Bray A, Ismay C, Chasnovski E, Baumer B, Cetinkaya-Rundel M. infer: tidy statistical inference. 2020. URL: https://CRAN.R-project.org/package=infer [accessed 2021-02-18]
23. Wickham H. ggplot2. Elegant graphics for data analysis. 2016. URL: https://link.springer.com/book/10.1007/978-3-319-24277-4 [accessed 2021-02-18]
24. Schmieding ML, Mörgeli R, Schmieding MAL, Feufel MA, Balzer F. Data set supplementing "Benchmarking triage capability of symptom checkers against that of medical laypersons: Survey study". Zenodo. 2021. URL: https://zenodo.org/record/4454538#.YJL3EqlKj0p [accessed 2021-05-05]
25. Mills B, Hill M, Buck J, Walter E, Howard K, Raisinger A, et al. What constitutes an emergency ambulance call? Australasian J Paramed 2019 Mar 22;16. [doi: 10.33151/ajp.16.626]
26. Age and sex composition in the United States. United States Census Bureau. 2019. URL: https://www.census.gov/content/census/en/data/tables/2019/demo/age-and-sex/2019-age-sex-composition.html [accessed 2020-08-10]
27. Meyer AND, Giardina TD, Spitzmueller C, Shahid U, Scott TMT, Singh H. Patient Perspectives on the Usefulness of an Artificial Intelligence-Assisted Symptom Checker: Cross-Sectional Survey Study. J Med Internet Res 2020 Jan 30;22(1):e14679 [FREE Full text] [doi: 10.2196/14679] [Medline: 32012052]
28. Semigran HL, Levine DM, Nundy S, Mehrotra A. Comparison of Physician and Computer Diagnostic Accuracy. JAMA Intern Med 2016 Dec 01;176(12):1860-1861. [doi: 10.1001/jamainternmed.2016.6001] [Medline: 27723877]

Abbreviations

HMS: Harvard Medical School
MTurk: Mechanical Turk
PDT: Pacific Daylight Time
SC: symptom checker

Edited by G Eysenbach; submitted 24.09.20; peer-reviewed by M Hill, E Berner, J Knitza; comments to author 04.10.20; revised version received 22.10.20; accepted 18.01.21; published 10.03.21.

Please cite as:
Schmieding ML, Mörgeli R, Schmieding MAL, Feufel MA, Balzer F
Benchmarking Triage Capability of Symptom Checkers Against That of Medical Laypersons: Survey Study
J Med Internet Res 2021;23(3):e24475
URL: https://www.jmir.org/2021/3/e24475
doi:10.2196/24475
PMID:33688845

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org